Firms, Productivity and Innovation in Low-Income Countries



Vanessa Schreiber

St. Catherine's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy in Economics

September 2020

Firms, Productivity and Innovation in Low-Income Countries

Vanessa Schreiber

St. Catherine's College
University of Oxford

A thesis submitted for the degree of Doctor of Philosophy in Economics

September 2020

This dissertation consists of three empirical essays on firms, productivity and innovation in low-income countries. Chapter 1 examines whether changing management practices to encourage bottom-up innovation leads to beneficial information flows from production workers on lower hierarchical levels to higher-level management with the power to implement ideas. In a three-month field experiment with 1,600 workers in a Bangladeshi garment factory, I find that providing structured peer-to-peer lunches for workers to discuss ideas significantly improves the quantity and relevance of ideas which are implementable in the firm, driven by changes in workers' ease of submissions and their aspirations. Additional encouragement from trained supervisors, however, comes at the cost of suppressing voice on critical topics about supervisors, which works through a perceived deteriorated voice climate. Chapter 2 (with Christopher Woodruff, Atonu Rabbani and Rocco Macchiavello) explores the impact of a low-cost group-based management training programme on productivity and working conditions in 25 Bangladeshi garment factories, with significantly positive effects on workers. Practices aimed at improving communication on the production floor are found to be more effective than technical and production-related practices, and lead to positive effects on human-resource related outcomes. Chapter 3 (with Ferrán Vega and Christopher Woodruff) analyses detailed line-level data in a Bangladeshi garment factory over a period of five years to examine the role of low-level autonomy in the process of adaptation to a negative productivity shock, caused by switching to a 'faster fashion' production model imposed by industry demands. We find that performance differences can be traced back to varying degrees of autonomy exercised by production managers in deviating from pre-determined plans.

Supervisors This thesis was supervised by Christopher Woodruff and Simon Quinn.

Total words This thesis contains approximately 88,165 words.

Statement of authorship Chapter 2 is joint work with Christopher Woodruff, Atonu Rabbani and Rocco Macchiavello. Chapter 3 is joint work with Ferrán Vega and Christopher Woodruff.

Acknowledgement of funding I gratefully acknowledge funding from the Doctoral Scholarship of the Konrad-Adenauer Foundation and the Final-Year Departmental Doctoral Bursary from Oxford's Department of Economics. I am thankful for generous financial support from the PEDL Special Exploratory Research Grant for PhD Students (Private Enterprise Development in Low-Income Countries).

Ethical approval Ethical approvals have been sought and obtained on the 26th of November 2019 from the Oxford Department of Economics' DREC with reference ECONCIA19-20-24 (Chapter 1), on the 30th of May 2014 from the University of Warwick Research Ethics Committee with reference 83/13-14 (Chapter 2), and on the 7th of August 2018 from the Oxford Department of International Development's Departmental Research Ethics Committee (DREC) with reference C1A_18_079 (Chapter 3).

To my family

Acknowledgements

2020 is not shaping up to be the year the world thought it would be. I also would never have thought I would have had to halt my field projects, cancel my research trip to Stanford and many more conferences. A challenging time to finish a PhD, but a time where I was, more than ever, reminded of the privilege I have had to work with such wonderful supervisors, peers, and the Oxford research community over the past three years, all of whom I thank deeply and unreservedly.

To my fantastic team of supervisors, who provided encouragement and guidance throughout my time at Oxford. To my primary supervisor, Christopher Woodruff: thank you for tirelessly supporting me, challenging me to strive for improvement, and always being a reference point to turn to for advice. I am deeply grateful for the wisdom and infinite knowledge you have imparted on me. I could not have asked for a better mentor and co-author. To my secondary supervisor, Simon Quinn: thank you for your guidance, and for being a role model in terms of your rigorous approach to research, your drive, and your thoughtfulness.

To inspiring academics I have had the privilege of learning from, the remaining faculty in the Center for the Study of African Economies (CSAE) and the Oxford Firms, Labour and Organisations group: thank you for laying a solid foundation of knowledge and for your kindness and feedback. More importantly, thank you for providing an academic 'home' for me in Oxford, and for being such a remarkably friendly research community. In the micro-theory group, to Meg Meyer and Vatsal Khandelwal: thank you for being sparring partners for my theoretical framework.

To my wonderful field team at Innovations for Poverty Action (IPA), Anaise, Arshiya, Awal, Busra, Jesmin, Marwan, Mithu, Nutan, Raju, Salman and Shankar: thank you for your outstanding field assistance, for much laughter during long hours on the road to the factories, and for going above and beyond during the outbreak of Covid-19. To Ishrat and her family: thank you for being the best remote family in Dhaka I could have ever prayed for, for your incredibly warm hospitality, and for making the stays in Dhaka so hassle free and enjoyable.

To Tatjana, the most caring and lovely housemate, and to Cécile and Lisa: thank you for our numerous talks and discussions about life, for your laughter, for making me feel at home at Oxford and for being a major source of inspiration! I will always cherish our friendship as a most precious gift. To my dear Econ crew, Ashley, Ben, Emma, Hannah, Katherine, Mahreen, Myriam, Myrto and Verena: thank

you for your friendship during every single difficult or joyful moment, your invaluable support, and for making work at the department so much fun. To my great motivator Lukas: I thank you wholeheartedly for always being so helpful. Without your friendly, caring and thoughtful presence, I would have not been able to get where I am. To the others that made my DPhil days so much brighter, Christian, Hannah, Margaryta, Nick, and Sangha: thank you for cheering me up. To Chris: thank you for helping me choose my supervisors. And to my co-author, Ferrán: thank you for being a great collaborator and a pleasure to work with.

To the others inside and outside of Oxford, who crossed paths with me in various ways and at different stages of my life, but all mean so much to me. My dear friends from the Oxford University Orchestra, Christine-Marié, Jannie and Poppy: thank you for sharing the love for music and for making rehearsals, concerts, and the Japan tour the most welcome distraction from my PhD. My dear Oxford friends, Arno, Ceren, Giulia, Katie, Marcel, Max and Sophie: thank you for contributing to making my time here worth remembering forever. My dear London friends, Charlotte, Dominik, Rodrigo and Sara: thank you for taking me out of the bubble. My dear Oxford friends from MPhil times, among others, Alejandro, Clara, Cynthia, Emily, Felipe, Fozia, Marcel, Martina, Rachita, Santi and Viviana - you are always close to my heart. My dear German/Swiss friends, Annika, Aylin, Caro, Catherine, Clemens, Conny, Fabi, Karsten, Kathi, Luisa, Sabine, Sahil, Sam, Yasmin and many others: thank you for always reminding me that home is waiting for me with a warm hug, and Mickey: thank you for your unwavering friendship despite the distance in space and time that often keep us apart.

My warmest thoughts go to my family to whom I dedicate this thesis, my parents, my brother, my grandparents and my wider family: thank you for your unconditional love and support in all of my pursuits, for encouraging me in my education, and for always backing my choices with your enthusiasm and trust. Throughout my life, and throughout all the challenges and frustrations that somehow seem to be a necessary complement of the doctoral studies, the genuine and constant support from you has really kept me going, and I owe you much more than words alone can describe. This thesis is also in memory of my grandparents who passed away during the writing of this thesis. And last, but not least, to Constantin: thank you for always pushing me to raise my ambitions, for keeping me grounded when Hogwarts was getting too much, and for cheering me along during the final steps of writing the thesis.

I gratefully acknowledge the financial support of the Konrad-Adenauer Foundation, the Oxford Department of Economics, and the Private Enterprise Development in Low-Income Countries (PEDL) programme.

Vanessa Schreiber
Oxford, September 2020

Contents

Introduction **1**

1 Fostering Innovation through Empowered Workers - Experimental Evidence from the Bangladeshi Garment Industry **7**
1.1 Introduction 8
1.2 Experimental design 14
1.3 Data collection and measurement 23
1.4 Main outcomes 37
1.5 Mechanisms 46
1.6 Conclusion 63

Appendices of Chapter 1 **67**
1.A Theoretical framework 68
1.B Determination of line location of workers 77
1.C Implementation details and documentation 78
1.D Sixth-step randomisation approach into lunch treatment 84
1.E Components of main outcomes 88
1.F Supplementary figures 98
1.G Supplementary tables: Main outcomes 103
1.H Supplementary tables: Mechanisms 108

2 Does Putting the Worker First Work? Evaluating a Worker-Centred Management Training Programme **117**
2.1 Introduction 118
2.2 Context in Bangladesh and sample selection 121
2.3 Experimental design 123
2.4 Data collection and measurement 129
2.5 Impact on worker outcomes 136
2.6 Impact on firm outcomes 145
2.7 Conclusion 152

Appendices of Chapter 2 **156**
2.A Description of the intervention 157
2.B Description of the data 164
2.C Empirical specifications 171
2.D Supplementary figures and tables: Worker outcomes 174
2.E Supplementary figures and tables: Firm outcomes 190

3 Does the Devil wear Zara? - The Impact of Fast Fashion and Low-level Autonomy in the Bangladeshi Garment Industry **201**
3.1 Introduction 202
3.2 Data 207
3.3 Qualitative description of trends over time 211
3.4 Empirical strategy 217
3.5 Results 220
3.6 Learning and autonomy within organisational entities 225
3.7 Conclusion 238

Appendices of Chapter 3 **242**
3.A Additional figures 247
3.B Additional tables 254

Conclusion **262**

References **265**

List of Figures

1.1 Overview of 2x3 design 16
1.2 Timeline 21
1.3 Total number of suggestions per round 38
1.4 Distribution of quality of submitted suggestions 40
1.5 Effects on content of suggestions by treatment groups 41
1.6 Daily effects on workforce index components 44
1.7 Illustration of heterogeneous treatment effects by experience and friendship 55
1.B.1 Data sources for line location of workers 77
1.C.2 SMS texts for supervisors 78
1.C.3 Suggestion paper 79
1.C.4 Template for line supervisors to track conduct of huddles 79
1.C.5 Template for evaluation committees to track choice of winning ideas 80
1.C.6 Sample display of winning suggestions on the floor's notice board . 80
1.D.7 Frequency of workers by number of times being mentioned as friend 85
1.E.8 Categorizations of suggestions 92
1.E.9 Secondary outcome variables 95
1.F.10 Distribution of total number of lines per worker 98
1.F.11 Distribution of maximum number of days per line per worker . . . 99
1.F.12 Effects on primary outcome variables 99
1.F.13 Average number of suggestions per round 100
1.F.14 Quality of submitted suggestions per round 101
1.F.15 Total number of submissions per worker 102
1.F.16 Repetition of categories by huddle treatment group 102

2.1 Illustration of effects for indices and sub-indices 141
2.A.1 Timeline 157
2.A.2 Training modules 159
2.A.3 Evaluation of practices 161
2.A.4 Chains of command 162
2.A.5 Assignment into batches 164

2.B.6 Definition of outcome variables . . . 165
2.B.7 Components of line operator indices . . . 166
2.B.8 Components of line supervisor/line chief indices . . . 167
2.B.9 Components of manager indices . . . 168
2.B.10 Attendance of key attendants by batch . . . 169
2.B.11 Attendance of top-level managers by department . . . 169
2.B.12 Attendance of middle- and low-level managers by department . . . 170
2.C.13 Illustration of empirical specifications: Worker outcomes . . . 171
2.C.14 Illustration of independent variables: Worker outcomes . . . 172
2.C.15 Illustration of empirical specifications: Firm outcomes . . . 173
2.C.16 Illustration of independent variables: Firm outcomes . . . 173
2.D.17 Dynamics over time for line operators . . . 178
2.D.18 Dynamics over time for line supervisors/line chiefs . . . 179
2.D.19 Managers' most important perceived changes in policies . . . 183
2.D.20 Managers' most important perceived changes in own job . . . 184

3.1 Efficiency over time . . . 212
3.2 Imposed industry demand . . . 213
3.3 Efficiency dispersion (2015, 2017 and 2019) . . . 214
3.4 Efficiency over time for same styles and all styles . . . 215
3.5 Comparison of effects of working knowledge . . . 225
3.6 Percentage of correctly planned styles by planning horizon . . . 227
3.7 Percentage of correctly planned styles by planning horizon per year 228
3.8 Ratio of number of produced and planned styles . . . 229
3.9 Percentage of correctly planned styles on any production line (within the same group) by planning horizon per year . . . 231
3.10 Production locations of correctly planned styles . . . 232
3.11 Efficiency costs in production and plan . . . 234
3.12 Delta in efficiency costs in production and plan across years . . . 235
3.13 Differences in efficiency costs per unit . . . 236
3.14 Decomposition of efficiency costs . . . 237
3.A.1 Alteration rates and rejection rates over time . . . 247
3.A.2 Imposed industry demand by unit . . . 248
3.A.3 Days between sewing start and sewing end date . . . 249
3.A.4 Days between sewing start and delivery date . . . 250
3.A.5 Percentage of correctly planned styles by unit . . . 251
3.A.6 Percentage of correctly planned styles on any line (within the same group) by unit . . . 252
3.A.7 Percentage of correctly planned styles on any line . . . 253

List of Tables

1.1 Balance table of primary outcome variables . . . 32
1.2 Balance table of worker characteristics and secondary outcome variables . . . 33
1.3 Effect of skill level on line movements . . . 35
1.4 Attrition . . . 36
1.5 Effects on primary outcome variables related to innovation . . . 37
1.6 Effects on primary outcome variables related to firm performance and worker well-being . . . 42
1.7 Effects on workers' perceived ease of idea generation, ease of submission and aspirations . . . 48
1.8 Heterogeneous treatment effects by workers' experience . . . 51
1.9 Interaction effects of workers' own and their peers' experience . . . 53
1.10 Heterogeneous treatment effects by supervisors' disposition . . . 57
1.11 Real satisfaction or unspoken dissatisfaction with supervisors . . . 59
1.12 Effects on communication practices and workplace culture . . . 62
1.C.1 Intervention dates . . . 78
1.C.2 Training curriculum . . . 81
1.C.3 Overview of huddles . . . 82
1.C.4 Overview of winning suggestions . . . 83
1.D.5 Overview of two randomisation samples . . . 84
1.D.6 Treatment assignments . . . 86
1.D.7 Number of exits before change of seating order . . . 87
1.D.8 Change of seating order . . . 87
1.E.9 Primary outcome variables . . . 88
1.E.10 Variables to measure heterogeneous effects . . . 91
1.G.11 LATE (IV) Effects on primary outcomes (firm performance and worker well-being) . . . 103
1.G.12 LATE (IV) Effects on primary outcomes (innovation) . . . 104
1.G.13 Effects on sub-components (firm performance and worker well-being) . . . 105

1.G.14 Effects on quality index components . . . 106
1.G.15 Effects on different specifications of variety . . . 106
1.G.16 Effects on the content of suggestions . . . 107
1.H.17 Learning over time . . . 108
1.H.18 Effects on secondary outcome components . . . 109
1.H.19 LATE (IV) Effects on secondary outcome components . . . 110
1.H.20 Heterogeneous treatment effects by workers' skills . . . 111
1.H.21 Heterogeneous treatment effects by workers' own gender and experience . . . 112
1.H.22 Heterogeneous treatment effects by experience and friendship . . . 113
1.H.23 Heterogeneous treatment effects by new seating partners . . . 114
1.H.24 Heterogeneous treatment effects by gender disposition and workers' experience . . . 115
1.H.25 Heterogeneous treatment effects on job satisfaction . . . 116

2.1 Examples of HR and production practices . . . 125
2.2 Summary statistics . . . 132
2.3 Balance table: Worker outcomes . . . 134
2.4 Balance table: Firm outcomes . . . 135
2.5 HR and PR index: All control lines . . . 139
2.6 Wellbeing and communication index: All control lines . . . 140
2.7 Spillover analysis . . . 145
2.8 Overall effects: All control lines . . . 147
2.9 Summary table of overall effects . . . 150
2.D.1 HR and PR index: Control lines on same floor . . . 174
2.D.2 Wellbeing and communication index: Control lines on same floor . . . 175
2.D.3 HR and PR index: Control lines on different floor . . . 176
2.D.4 Wellbeing and communication index: Control lines on different floor 177
2.D.5 Sub-Indices for line operators: All control lines . . . 180
2.D.6 Sub-Indices for line supervisors and line chiefs: All control lines . . . 181
2.D.7 Sub-Indices for managers: All control lines . . . 182
2.D.8 Buddy system: Operators . . . 185
2.D.9 One-step leave policy: Operators . . . 185
2.D.10 Operators' feelings: Operators . . . 186
2.D.11 Communication practices: Operators . . . 187
2.D.12 Supervisor/line chief knowledge: Operators . . . 188
2.D.13 Quality practices: Line supervisors/line chiefs . . . 188
2.D.14 Buddy system: Managers . . . 189
2.D.15 Quality practices: Managers . . . 189
2.E.16 Overall effects: All control lines . . . 190
2.E.17 Overall effects: Control lines on different floors . . . 191

2.E.18 Spillover effects: Adjacent lines in all factories . . . 192
2.E.19 Spillover effects: Adjacent lines . . . 193
2.E.20 Spillover effects: Adjacent and other lines . . . 194
2.E.21 Spillover effects: Control floors . . . 195
2.E.22 Spillover effects: All floors . . . 196
2.E.23 Robustness check based on floor information: Summary table . . . 197
2.E.24 Robustness check based on floor information: All control lines . . . 198
2.E.25 Robustness check based on line information: Summary table . . . 199
2.E.26 Robustness check based on line information: All control lines . . . 200

3.1 Summary statistics . . . 209
3.2 Effects on efficiency using all styles and common styles . . . 216
3.3 Overview of variables related to production . . . 217
3.4 Standard specification . . . 223
3..1 Match rates from style library . . . 244
3..2 Differences between matched and non-matched observations . . . 245
3..3 Match rates by year and unit . . . 246
3.B.4 Overview of variables related to production . . . 254
3.B.5 Standard specification with line fixed effects . . . 255
3.B.6 Standard specification with line fixed effects and clustered standard errors . . . 256
3.B.7 Standard specification by year . . . 257
3.B.8 Selected units with style fixed effects . . . 258
3.B.9 Effects on efficiency using all styles and common styles (full table) . 259
3.B.10 Standard specification for industry . . . 260
3.B.11 Standard specification by unit . . . 261

Introduction

This thesis presents research on two topics relating to organisational practices in manufacturing firms in low-income countries, with a particular focus on the Bangladeshi garment industry. The first topic is how managerial practices can foster bottom-up innovation. Chapter 1 aims to provide experimental evidence on this topic by studying the causes and consequences of fostered upward information flows from production workers to managers. The second topic is about determinants of performance differences within and across firms, and how such differences can be reduced in a cost-effective way. Chapter 2 and 3 present research on this topic by experimentally testing effects of a low-cost management training on firm performance and worker well-being, and by non-experimentally analysing the role of learning processes and autonomy, respectively.

Motivation In 1887, Francis Walker, the first president of the American Economic Association, already recognized the importance of managers – "men who have high power of administration" – as sources of performance differences across firms, and supplemented the trinity of land, capital and labour with the new entity *management* (Bloom et al., 2014):

> *"In works controlled by men who have a high power of administration and a marked degree of executive ability, where everything goes smoothly and swiftly forward to its end, where emergencies are long foreseen and unfavourable contingencies are carefully guarded against, where no steps have to be retraced, and where nothing ever comes out wrong end foremost, there is much less nervous and muscular wear and tear than in works under inferior management. [...] Side by side, in the same business, with equal command of capital, with equal opportunities, one man is gradually sinking a fortune, while another is doubling or trebling his accumulations."*
>
> Francis Walker, *The Quaterly Journal of Economics*, 1887

Despite the fact that the idea of management as a driver of productivity dates back to the 19th century, economists have only recently become increasingly aware of

the importance of organisational practices on firm performance. Historically, importing foreign know-how has widely been acknowledged as a driver of growth, but the focus was directed at tangible know-how, such as modern sewing machines or other hard technology, mainly because of its ease of measurability and procurement. Over the last decade, more policy effort and money has been dedicated to intangible know-how, such as management practices. Economists have started to systematically measure management practices[1] and discovered the high potential of intangible capital in yielding returns with lower levels of investment – as *Bloomberg* put it, "Management consultants might be the best foreign aid" (Smith, 2019). For example, managerial capital has been found to account for large performance differences within and across countries (Bloom et al., 2017). At the micro level, several studies find a large causal role for management practices in increasing productivity (Bloom et al., 2013; Bruhn et al., 2018; Giorcelli, 2019). Particularly in increasingly competitive industries, the adoption of better management practices can be a key success factor. The Bangladeshi garment industry serves as an ideal setting to study this. It contributes one-eighth of the country's economic output and accounts for more than 80% of export earnings. In 2019/20, exports of ready-made garments generated foreign-exchange earnings of $27.9bn (BGMEA, 2020). The industry has generated around four million jobs for Bangladeshis, most of whom are women, have eight or fewer years of schooling and come from lower-income backgrounds. The rapid growth of the sector has fuelled annual per-capita income growth of around six percent for the past two decades (The World Bank, 2020). This growth increased wages in the broader economy. Though instrumental to national development, garment factories are viewed as inefficient in comparison to many countries with which Bangladesh competes. In spite of this relative inefficiency, especially for lower-end products (e.g., t-shirts) that represent the bulk of exports, low labour costs have kept the sector competitive internationally. However, the sector faces various on-going struggles and risks, jeopardizing its growth. First, the industry is characterised by poor working conditions and workplace safety, as shown by the collapse of the Rana Plaza building in 2013 causing

[1]The World Management Survey (WMS) was set up to systematically measure management practices across countries and sectors. The survey was developed to explain the large and persistent total factor productivity (TFP) differences across firms and countries, first described in Bloom and Van Reenen (2007) and extended in Bloom et al. (2017).

the death of more than 1,100 garment workers, or the reports of workplace harassment.[2] As buyers become increasingly prompted by the international stage to comply with labour standards, the industry faces additional pressure. Withdrawals of preferential trade benefits, such as in 2013 from the US in consequence of the Rana Plaza incident, further contribute to decreasing competitiveness of the industry (The Economist Intelligence Unit, 2019). Second, ongoing struggles over worker safety and well-being puts the industry at the high risk of protests and demonstrations from factory workers, additionally decreasing efficiency. Third, the industry faces increasing automation. While initial costs to set up automated factories are still high, automation will pose a serious threat to the sector's future with technological advancements, rising wages in the Banlgadeshi economy, and nearshoring becoming economically viable (McKinsey & Company, 2018). Overall, adaptations of the workers' skill sets, movements to more complex products, and efficiency gains will be key to the industry's continued prosperity and to remain competitive in the export market for garments. A better understanding of how Bangladeshi garment factories achieve productivity improvements in conjunction with improved worker well-being is becoming increasingly important.

Outline This thesis has three core chapters and uses a combination of experimental data (Chapter 1 and 2) and non-experimental data (Chapter 3) to identify key issues affecting productivity and worker well-being in the Bangladeshi garment sector. In Chapter 1, I analyse barriers to bottom-up innovation, and how upward flows of ideas from production workers to managers can be fostered through managerial practices. There is relatively little evidence on the effect of management practices in developing countries which foster employee decision-making. While the involvement of workers into firm's decision-making processes seems to be rare in relatively hierarchical and low-trust business environments, including the Bangladeshi garment industry, the potential benefits could be large. For example, Toyota successfully deploys suggestion systems as a form of employee participa-

[2]The working conditions and workplace safety are even poorer in the majority of other factories which are not part of this thesis, and in their subcontractors, as shown, for example, by the regular reports of child labour or a number of factory fires (Paton, 2020).

tion with yearly cost savings of potentially $70 million.[3] I set up suggestion boxes for 1,600 workers in a Bangladeshi garment factory, which offered awards for the top three ideas each month. I also implemented two cross-randomised managerial interventions through a randomised controlled trial. I show that managerial interventions in the form of providing structured peer-to-peer lunches can improve the quality and relevance of employees' ideas implementable in the firm, and simultaneously lead to higher job satisfaction. On the contrary, managerial interventions in the form of encouragement from trained supervisors do not serve as an accelerator to make peer interactions during lunch more efficient, and even come at the cost of steering workers away from topics they would otherwise raise, such as supervisors' skills. I claim that the first intervention is successful because it makes it easier for workers to submit ideas and increases their aspirations. Further, I find evidence of a peer-mentor effect: less-experienced workers benefit from being seated opposite a more-experienced friend during lunches. Motivated by this evidence, I build a theoretical framework to consider possible reasons behind better idea generation with co-workers who are friends or non-friends than alone.

In Chapter 2, written with Christopher Woodruff, Atonu Rabbani and Rocco Macchiavello, we analyse whether more cost-effective management interventions can improve both productivity and worker well-being at the same time. The renowned consulting firm Accenture delivered the services for a successful experiment in India (Bloom et al., 2013), raising the question whether cheaper local consultancy firms can deliver effective interventions at a cost lower than global consultancy firms can. To provide an answer to that question, we conduct a low-cost group-based management training programme in 25 Bangladeshi garment factories in the form of a randomised controlled trial. We find that the programme has significant positive effects on workers. The strongest effects are on human-resource related outcomes, leading to decreases in exit rates, turnover and absenteeism. Practices aimed at improving communication on the production floor seem to have been more effective than technical and production-related practices. However, most of the effects disappear within six months of treatment. We see some evidence of spillovers of practices, particularly to production lines on the same

[3] Assumption: $100 cost saving per idea with 700,000 implemented suggestions per year at Toyota (Sting et al., 2019)

floor as the treatment line, but the spillovers do not appear to explain the fade out of the treatment effects.

In Chapter 3, written with Ferrán Vega and Christopher Woodruff, we shed light on drivers of productivity and differences in productivity within manufacturing firms. We use very detailed firm-level data from 118 production lines in a Bangladeshi garment factory to examine the role of low-level autonomy in the process of adaptation to a negative productivity shock. We begin by qualitatively describing trends in the data, in particular (i) a large decrease in efficiency, caused by switching to a 'faster fashion' production model imposed by industry demand, and (ii) an impressive, but uneven, recovery despite unchanged conditions. We then examine those trends in more detail. First, our data suggests that the recovery was triggered by learning processes, for example by building 'working knowledge' at the level of the individual item (style) being produced. This increases for each additional day a style is produced within a given production spell. Second, our data suggests that performance differences across units can be traced back to varying degrees of autonomy exercised by production managers in deviating from pre-determined plans. After comparing the efficiency costs in the production data with those embedded into the plans, we find that the change in the factory's production model undermined the planning departments' ability to generate adequate production plans. We argue that deviating from plans to re-organise production within units' organisational groups was fundamental to keep up with a high-paced industry.

Contribution As a unified piece of work, this thesis contributes to the literature on firms, innovation and productivity in low-income countries by providing new experimental evidence on management interventions and non-experimental evidence on determinants of productivity differences.

Chapter 1 provides four key contributions. First, I contribute to the growing literature on the impact of management practices on firm-level outcomes by providing evidence on the success of managerial interventions in fostering knowledge flows across hierarchies. Second, I contribute to the literature on worker participation in firms' decision-making processes by investigating the causal impact of

increasing workers' voice. Third, I contribute to the literature on the role of hierarchies in processing and communicating information by showing that involving middle managers might suppress worker voice on certain topics. Fourth, I touch upon the literature on worker voice in management science and organisational psychology, showing how managerial interventions affect workers' perceived psychological safety, and their aspirations.

Chapter 2 provides a number of contributions on the emerging literature on experimental evaluation of training programmes and effects of group-based training programmes on firm performance. The first and main contribution is to evaluate the causal impact of low-cost management training on both performance and worker well-being, differentiating between productivity- and human resource-related training modules. The second key contribution is to provide a detailed spillover analysis using rich administrative and survey data.

Chapter 3 provides three primary contributions. First, we contribute to the non-experimental literature on determinants of productivity differences by providing a detailed decomposition of production into its internal and external determinants and estimate their effects on productivity- and quality-related outcomes. Second, we contribute to the literature on learning and adaption within and across firms, by investigating learning effects and autonomy of production units. Third, while our contribution is empirical, our work links to theoretical papers on organisational failures and the importance of delegation residing at lower levels of the organisation.

CHAPTER 1

Fostering Innovation through Empowered Workers - Experimental Evidence from the Bangladeshi Garment Industry

Abstract Informational constraints are often a crucial barrier to adoption of innovations in firms. How can the culture in firms be changed to overcome this barrier and to increase upward flows of ideas? I set up suggestion boxes for 1,600 workers in a Bangladeshi garment factory and test the efficacy of two cross-randomized voice-enhancing managerial interventions through a randomized controlled trial. First, I randomly assign workers to have time set aside with co-workers to discuss ideas during free lunches. Second, I assign supervisors on randomly-selected production lines to be trained on communication methods and to encourage workers to develop ideas. I find that the first intervention fosters upward flows of ideas, and simultaneously leads to improved worker well-being. The second intervention, however, does not serve as an accelerator to make peer interactions more efficient and even comes at the cost of undermining workers' critical voice about supervisors. The evidence suggests that the former result can be explained by changes in workers' aspirations and perceived ease of submission, whereas the latter result is mainly driven by a perceived deterioration in the voice climate. Further, I find that specific seating arrangements opposite experienced friends during lunches can be particularly beneficial for inexperienced workers.

1.1 Introduction

"Workers on the production line might observe quality-defect problems that become apparent only on the shop floor, or notice a machine that is sometimes idle, or a surplus stock of raw materials that could be used." (Hayek, 1945)

As Hayek stated in 1945, knowledge valuable to an organisation is acquired by people at all levels of the organisation including the lowest, and often consists of information about things that are seemingly trivial. Hayek's view reflects a problem which is still prevalent in many organisations today: Information that would be helpful for the firm is not brought up the hierarchy. In this paper, I examine whether changing management practices to encourage bottom-up innovation leads to beneficial information flows from production workers on lower hierarchical levels to higher-level management with the power to implement ideas. I provide the first experimental evidence on impediments to idea-driven voice in manufacturing firms low-income countries.

What limits communication of ideas from production workers to managers in hierarchical organisations? On the one hand, workers may face costs of effort and time to develop ideas, arising from, for example, inabilities to generate suggestions in an organisational language that managers can understand and reward (Weber and Camerer, 2003) or opportunity costs of time (Battiston and Kirchmaier, 2018). On the other hand, workers may face social costs that limit the diffusion of information, arising from e.g. reputational concerns (Bursztyn and Jensen, 2017), fear of signaling incompetence (Edmondson, 1999), or class differences/altruism towards in-group co-workers (Hjort, 2014). I design experiments to assess the effectiveness of voice-enhancing managerial interventions that target costs of effort and time, social costs, and the combination of both of these. In doing so, I set up a suggestion system in a Bangladeshi garment factory on six production floors with a total of 1,588 workers and 51 production lines. The system offers awards for the top three ideas each month and displays the winners' names for recognition. In this setting, I conduct a randomised controlled trial (RCT) with two nested experiments between December 2019 and March 2020, using a cross-randomised design. The first experiment ("lunch treatment"), targets costs of effort and time. I follow Sandvik

et al. (2020) in randomly assigning workers to have time set aside with co-workers over a free lunch once per month. With the assistance of external facilitators, workers were advised to jointly discuss suggestions and fill out suggestion forms during lunchtime. For fairness reasons, workers in the control group receive a free take-away lunch and a suggestion form, but without structured opportunities to generate suggestions in the form of fixed time set aside with external facilitators. The second experiment ("huddle treatment") targets social costs. Supervisors on randomly selected production lines are assigned to complete a two-hour training program on communication methods and to encourage worker participation in the suggestion system through monthly line huddles, i.e. motivational meetings with workers lasting three to five minutes.

In this paper, I study a comprehensive set of worker and firm outcomes drawing upon a rich set of three data sources: factory data (production, quality and personnel data), intervention data (submitted suggestions) and survey data (baseline and endline surveys with workers and supervisors). I estimate the short-run impact of the interventions on two groups of outcomes: idea-driven voice on the one hand and worker well-being and firm performance on the other hand. From the workers' and firm's perspective, both sets of outcomes are important to analyse. First, giving workers a voice through encouragement of suggestions might make them feel more listened to and cared for. This empowerment might directly affect worker well-being and/or indirectly affect firm performance through higher motivation and effort of workers in the short-run regardless of the quality of suggestions. Second, collecting additional suggestions might be worthwhile in the short-run if their implementation has the potential to improve firm performance in the long-run, in line with the Japanese production philosophy *Kaizen*.[1]

Related to idea-driven voice, I find that the lunch intervention significantly increases the quantity and relevance of suggestions that are implementable in the firm. In fact, workers not assigned to lunches hardly submit any suggestions at all (48 suggestions from the control group compared to 1809 of the treatment group). However, the quality of the suggestions submitted by the control group is

[1]The Japanese concept of *Kaizen* (continuous improvement) covers different management techniques, in which everyone, workers and managers alike, are requested to consider continuous improvements as their permanent duty (Aoki, 1990).

20% higher, where quality is an index of six criteria evaluated by an industry expert. This result suggests a positive self-selection of high-quality submissions from workers in the non-lunch group. On the contrary, workers in the lunch group submit suggestions of much larger breadth, covering diverse topics. The huddle treatment has no effects on the quantity, quality or variety of suggestions, where variety is measured based on a count of new categories over time, and for robustness also on a latent semantic analysis approach (LSA) comparing word vectors. However, the huddle treatment leads to a reduced number of suggestions on critical topics about supervisors' skills and behaviours.

Related to worker well-being and firm performance, I find that the lunch intervention improves worker well-being in the form of higher job satisfaction at the time of the endline survey, but has no short-run effect on firm performance. However, I observe lower absenteeism on assigned lunch treatment days. This suggests that workers enjoy being involved in idea generation and receiving a free lunch gift from the firm. The huddle treatment has neither an effect on worker well-being and firm performance on its own nor in combination with the lunch treatment.

Overall, the huddle treatment is not beneficial. In contrast, the lunch treatment has clear benefits. I provide three explanations in favor of trading off quality against quantity, i.e. to get more suggestions of varying quality from the lunch group rather than fewer suggestions of high quality from the non-lunch group. First, the lunch treatment makes workers better off regardless of the quality of submitted suggestions, which could eventually lead to reduced turnover in the long-run. Second, the lunch treatment is a cost-effective intervention in improving worker well-being; it is not costly to evaluate suggestions, and 88% of the suggestions judged to be best (blind to treatment) were from the lunch group. Third, the lunch treatment enables the firm to get a feeling for what workers think and care about through the larger breadth of submitted topics. Implementing suggestions on those topics could eventually lead to productivity improvements in the long-run.

Next, I examine the potential mechanisms by which the two managerial interventions have an impact. First, I find that voice at the extensive margin is driven

by changes in workers' perceived ease of submission and their aspirations. Treated workers in the lunch group report that they feel more incentivised and comfortable to submit suggestions, and they are also more optimistic about winning and having their idea implemented. Second, I find that voice at the intensive margin is driven by workers' individual abilities. Workers with an above-median experience index at baseline submit more and better suggestions.[2] Further, I find evidence of a peer-mentor effect: less-experienced workers benefit from being seated opposite a more-experienced friend during lunches.[3] The effect of friendship is identified because half of the workers in the lunch group were randomly assigned to be seated opposite a friend. Third, I find that the huddle treatment comes at the cost of undermining workers' critical voice about their supervisors measured by a perceived deterioration in the voice climate, consistent with an effect defined in the literature as low 'psychological safety' (Edmondson, 1999; Edmondson and Lei, 2014). Workers exposed to treated supervisors report in several questions that they feel less secure about sharing their opinions.[4] Fourth, I find that job satisfaction is mostly driven by the act of exerting voice rather than perceived changes in the actual working conditions, consistent with the findings of Adhvaryu et al. (2019).[5]

This chapter makes four primary contributions. First, there is a growing literature on the impact of management practices on firm-level outcomes, with significant evidence that productivity differences can be causally explained by differences in management practices (Bandiera et al., 2020; Bloom et al., 2013, 2018;

[2]The experience index is an Anderson (2008) index of the number of factories worked at, grade, tenure in the present factory, tenure in the industry and the educational level. I find similar results using the above-median number of processes that a worker can perform, measured in skills assessments.

[3]These results relate to those in Hahn et al. (2019) who find that the individual performance of lower-ability female children in Bangladesh improves once they study with peer friends. Similarly, Sandvik et al. (2020) find that being paired with a more-experienced co-worker during structured peer-to-peer lunches increases average sales revenue.

[4]Focus group discussions with supervisors delivered qualitative evidence that supervisors are worried about workers complaining about them to top managers, suggesting that supervisors could have indeed steered workers away from raising certain topics in line with Friebel and Raith's theory on skip-level communication (Friebel and Raith, 2004).

[5]Adhvaryu et al. (2019) show that giving randomly selected workers the opportunity to participate in a feedback survey after a disappointing wage hike reduced turnover. As participants were told that the firm would not receive the feedback provided, they conclude that the act of being asked for one's opinion is driving the result rather than possible changes in firm policy or managers' behaviour.

Bruhn et al., 2018; Higuchi et al., 2017; Iacovone et al., 2018). The most closely related paper is the recent work by Sandvik et al. (2020) showing that interventions that provide structured opportunities for workers to initiate conversations with peers foster knowledge flows among co-workers on each other sales' techniques. By contrast, they find that monetary incentive systems based on joint output gains have no effects. While Sandvik et al. (2020) intervene with workers on the same hierarchical level, I focus on the effect of interactions between peers on bottom-up knowledge flows from production workers to higher-level managers. Plausibly, in my context there is much more to be learned from bottom-up knowledge flows than from peer learning due to the highly-specialised and manual nature of each workers' tasks, such as sewing certain processes. Unlike in Sandvik et al. (2020), idea generation is not achieved by efficiently connecting knowledge seekers and providers. Instead it requires complementing different pieces of information and conceptualizing combinations of knowledge, as the solution might be unclear to both knowledge seekers and providers.

Second, this paper contributes to the literature on worker participation in firms' decision-making processes. Grant (1996) develops a knowledge-based theory of the firm, which assumes that knowledge resides within each individual. In his model, a participatory culture that allows knowledge transfer can increase both worker well-being and efficiency.[6] Several studies provide empirical evidence on the success of employee suggestion systems (Gibbs et al., 2017), and more generally, the positive impact of increased worker voice in firms' decision-making processes on economic performance and workers' welfare. Examples include worker representation in corporate boards (Jäger et al., 2020), worker-manager safety committees (Boudreau, 2020), manager evaluations (Cai and Wang, 2020) or voice in the realm of unsatisfactory work conditions (Cottini et al., 2011) or disappointing wage hikes (Adhvaryu et al., 2019). Existing studies in low-income countries have not looked at the effects of giving workers a voice in the realm of improvements

[6]His theory closely corresponds to Aoki's 1990 analysis of the Japanese corporation where production decisions are decentralised to workers. The Japanese concept of *Kaizen* (continuous improvement) covers management techniques such as quality circles, total quality management, kanban, or just-in-time. Another way to put their concept into practice are suggestion systems to access and utilize individual's knowledge located at low levels of the organisation. For example, large industrial companies such as Toyota have successfully implemented suggestion systems. Sting et al. (2019) observe that more than 700,000 suggestions are implemented at Toyota per year.

in both productivity and working conditions, and effective managerial methods in doing so.

Third, I contribute to the literature on the role of hierarchies in processing and communicating information. While hierarchies are beneficial for the coordination of activities (Bolton and Dewatripont, 1994), communication across hierarchies can be costly. Reasons for these costs have been extensively analysed theoretically. For example, costs may arise due to limits to people's ability to transmit, receive or process information (Geanakoplos and Milgrom, 1991), monitoring of subordinates (Qian, 1994) or strategic use of information (McAfee and McMillan, 1995). More applicable to the context of this paper, communication costs can also arise from different organisational languages used by a principal and his agents, and a lack of a rich enough shared vocabulary (Arrow, 1974; Melumad et al., 1992; Weber and Camerer, 2003; Prat et al., 2007). Empirical evidence on barriers to communication and how to overcome those is limited, in particular on communication across hierarchies. Previous literature analyses, for example, incentive alignments between employees and owners (Atkin et al., 2017), adoption of innovative HR management practices (Gant et al., 2002) or co-location of workers (Battiston and Kirchmaier, 2018) to overcome informational constraints within and across hierarchies in firms. I add to this literature by showing that peer-to-peer interactions enable production workers to conceptualize and sharpen their messages in an 'organisational language' which top managers can understand and reward.[7]

Fourth, I touch upon the literature on worker voice in management science and organisational psychology which analyses the conditions under which individuals feel empowered to speak up.[8] Many studies on worker voice[9] find that leader

[7]Another aspect to which my paper relates to is the effect of 'skip-level' communication, i.e. communication between subordinates and top management. Middle managers can have strategic concerns when communication between their subordinates and superiors is too open (Baird, 1977; Lillico, 1972). In fact, Friebel and Raith (2004) suggest in their theoretical model that skip-level communication may induce superiors to deliberately recruit unproductive subordinates, or abuse their personnel authority in other ways, to protect themselves. I empirically show that the involvement of supervisors – who simultaneously act as principals (to workers) and agents (to higher-level managers) – comes at the cost of steering workers away from topics which might involve negative messages about them, in particular their skills and behaviour such as shouting or harassment.

[8]See for example Coffman (2014) on idea contribution in areas outside of a gender's domain, or Rigtering et al. (2019) on managerial framing for suggestion systems.

[9]See the literature overview by Morrison (2014) and Burris et al. (2018).

behaviour plays an important role in motivating employees to speak up by reducing the fear of repercussions. Similarly, workers' perceived psychological safety is found to be a key antecedent of risk-taking behaviour such as suggesting ideas (Edmondson, 1999; Edmondson and Lei, 2014). Williams and Yecalo-Tecle (2019) find that hostility by supervisors to new ideas from subordinates is an overwhelming constraint on voicing ideas for public servants. I further contribute by finding empirical evidence on the effect of managerial interventions in changing workers' perceived psychological safety (which I call 'voice climate'), and aspirations.

The paper proceeds as follows. In Section 1.2, I explain the experimental design of the randomised controlled trial. In Section 1.3, I discuss the data collected, the outcome variables and the empirical specification. In Section 1.4, I report the main impacts on firm performance and worker well-being, and idea-driven voice. Section 1.5 presents the results on mechanisms, and Section 1.6 concludes.

1.2 Experimental design

1.2.1 Setting

The study was conducted in a garment factory located in Gazipur, north of Dhaka in Bangladesh. The factory exports to large international buyers, one of whom requested the factory to set up suggestion boxes on each of the factory's sewing floors as a social compliance requirement. The factory agreed to collaborate on research to analyse how the suggestion submission process can best be facilitated. As a starting point, the factory placed suggestion boxes at easily accessible places for all workers on each floor. The factory made suggestion templates and pens available and requested that workers submit suggestions about improvements of both productivity and working conditions. The suggestion template is shown in Figure 1.C.3. It contains pre-specified areas of improvement: Absenteeism, defects, efficiency, machines, motions/waiting time, production flows/layout, safety, shipment, skills, wastage, work environment or other. Furthermore, the factory introduced prizes for the best three ideas per floor per month of 500 BDT (approximately 6% of a helper's and 5% of an operator's monthly salary), 300 BDT and 100 BDT. Financial rewards were set in place to directly share with workers the payoff of improvements they create and incentivise them to exert effort. Besides, floor

managers were asked to introduce the suggestion box to all workers on the floor via the public announcement system. In this setting, the suggestion box remained empty.[10] This suggests that knowledge flows across hierarchies are challenging in low-trust environments such as in Bangladeshi garment factories. "Nobody gives importance to the floor girls and they [supervisors/managers] won't even listen to us." a female Bangladeshi garment worker on the production floor said.[11] After testing different interventions to increase the number of submissions on one pilot floor, the factory agreed to roll out two managerial interventions on 51 production lines on six production floors. The study population was composed of 1,588 workers who were registered on those production lines at the program start date.

1.2.2 The interventions

I conducted two cross-randomised managerial interventions, which were designed to shed light on different impediments to bottom-up innovation: Costs of effort and time, social costs, and the combination of both of these. On the one hand, workers may face costs of effort and time, arising from, for example, inabilities to generate suggestions in an organisational language that managers can understand and reward (Weber and Camerer, 2003) or opportunity costs of time (Battiston and Kirchmaier, 2018). On the other hand, workers may face social costs, arising from e.g. reputational concerns (Bursztyn and Jensen, 2017), fear of signaling incompetence (Edmondson, 1999), or class differences/altruism towards in-group co-workers (Hjort, 2014).

The first managerial intervention ("lunch treatment"), targets costs of effort and time. It was implemented at the worker level (1,588 workers) and consisted of structured peer-to-peer lunches. To understand mechanisms, workers were randomly assigned to be seated opposite friends or random workers (called non-friends[12] throughout this chapter) during lunch. The control group received a free take-away lunch for fairness reasons. The second experiment ("huddle treatment")

[10] Even with more substantial monetary rewards in place, which were twice as large (approx. 12% of helpers' and 10% of operators' monthly salary), workers did not submit suggestions.

[11] Focus group discussion in one Bangladeshi garment factory with randomly selected workers, 04.11.2019.

[12] A non-friend is not defined as someone a worker does not like, but as someone a worker does not know.

targets social costs. It was implemented at the production line level (51 production lines) and consisted of a supervisor training program and line huddles with workers. Workers working on production lines assigned to the second managerial intervention were exposed to line huddles from trained supervisors, regardless of whether they were assigned to the first managerial intervention or not. The cross-randomised design allows me to categorize workers into one of the six groups of a 2x3 matrix as depicted in Figure 1.1. For simplicity, I refer to workers assigned to the first managerial intervention as the "lunch group" and workers exposed to the second managerial intervention as the "huddle group". A more detailed description of each intervention is described in the following subsections.

		Intervention 1: Peer-to-peer lunches (worker-level)		
		Control	**Treated** ("lunch group")	
			opposite friend	*opposite non-friend*
Intervention 2: Supervisor training and line huddles (line-level)	Control	Take-away lunch	Seated lunch opposite friend	Seated lunch opposite non-friend
	Treated ("huddle group")	Take-away lunch + Line huddle	Seated lunch opposite friend + Line huddle	Seated lunch opposite non-friend + Line huddle

Figure 1.1: Overview of 2x3 design

Notes: Figure 1.1 shows the cross-randomised experimental design. Workers are directly randomly assigned to the first managerial intervention, structured peer-to-peer-lunches, either to be seated opposite a friend or a non-friend. Depending on which production line they work at, they are indirectly assigned to the second managerial intervention, line huddles. Thus, workers can be clustered into on of the six buckets in the 2x3 matrix.

Managerial intervention 1: Lunches

The idea behind the first managerial intervention is to test whether separating out time for workers to engage in discussions with co-workers about suggestions under the help of external facilitators improves the efficacy of idea generation. The intervention reduced opportunity costs of time (Battiston and Kirchmaier, 2018),

arising from the fact that every second spent thinking about and writing down suggestion cannot be devoted to other activities. I followed Sandvik et al. (2020) in providing free monthly lunches to a random subset of workers. The factory provided space and tables so that I could arrange for workers to interact with co-workers. Simultaneously, an additional feature of this managerial practice was the provision of time with external facilitators. External facilitators handed out suggestion papers and pens to each worker and explained how to fill out the template. Workers were able to ask questions to external facilitators during lunch. Suggestions could be handed over to external facilitators at the end of the lunch, or submitted any time into the suggestion box on the production floor. Workers were seated in the cafeteria on long tables with up to 18 seats. For logistical reasons, workers were partitioned into four sub-groups based on their pre-assigned lunch hours. Every week, a different sub-group received lunch (in two cohorts from 12:00 to 12:45pm, and from 1:00 to 1:45pm), such that each sub-group received one lunch per month over the course of three months. Workers were informed about their assignment to a lunch session two to three days in advance as they otherwise often went home for lunch.[13]

For fairness reasons, the control group received a take-away lunch which was handed out together with a suggestion paper and a short verbal instruction. However, the control group was exposed to neither the natural opportunity to interact with peers as part of a fixed seating arrangement nor to assistance from external facilitators. Workers in the control group were free to eat lunch wherever and whenever they wanted, e.g. inside the factory or at home. Most took the lunch home for dinner or for their families and ate lunch as they usually would. They received a short instruction about how to fill out the suggestion paper, which was shorter in length than the one treated lunch workers received.

To measure the importance of trust in discussions, I randomised those receiving lunch to be seated opposite someone they identify as a friend or another randomly selected worker during lunch. A friend is defined as a co-worker on the production floor with whom a worker talks to most about work-related topics, and hence, can also trust to talk about suggestions. The baseline survey asked workers to

[13]More than 25% of the workers have lunches inside the factory less than every day, as recorded in the baseline survey. 15% of the workers never have lunch inside the factory.

name up to two co-workers they talk to most in the baseline survey. Workers were explicitly not asked to name 'friends', as many workers found it difficult to grasp the meaning of friendship in the pilot survey. In fact, less than 1% only talk about private topics and more than 80% only talk about work-related issues with the person they talk to most. As some workers were able to coordinate their answers in advance, and some were not, I impose symmetry on all friendships for simplicity.[14] After two lunches, I re-randomised the seating of a random sample of one-third of workers, who were moved to be seated opposite someone other than their friends. This allows me to test whether being seated next to someone new after two lunches stimulates new ideas.

Managerial intervention 2: Huddles

The second managerial intervention is at the production line level. Supervisors working on randomly selected production lines were assigned to take part at a 2-hour training program on worker communication methods at the start of the intervention and then instructed to conduct short motivational meetings of two to five minutes with all workers on their line (''line huddles''), explicitly encouraging the workers to submit suggestions. The intervention was designed to test whether being encouraged by supervisors serves as an additional accelerator to making the first managerial intervention more efficient.

The supervisors received training on how to conduct the huddles. The training focused on worker communication methods, with both a motivational and technical part. In the motivational part, supervisors learned about the new suggestion system and their roles and responsibilities during the suggestion process. They were taught about how to maintain good relationships with workers, how to encourage workers to submit suggestions (through on-the-job encouragement and line huddles) and how to give feedback. In the technical part, supervisors were asked to share their experiences with line huddles, learned about how to generate solution-oriented suggestions, and were given insights about possible areas of improvement. The curriculum is shown in Table 1.C.2.

[14] If A is mentioned as a friend of B, then B is also classified as a friend of A (even if A did not mention B as a friend).

To remind supervisors to conduct line huddles during the intervention period, I sent each treated supervisor three SMS reminders (see text in Figure 1.C.2) with the request to conduct a huddle within the next two to three days. After conducting a huddle, supervisors filled out a template like the one shown in Figure 1.C.4 and submitted it to their Industrial Engineering (IE) manager for approval. The template was introduced as a commitment device and to track whether huddles actually took place. The research team prompted IE managers to ensure that supervisors conducted the huddle and did not just fill out the template with fake information.

Evaluation process

Over the course of the intervention, the research team collected suggestions during lunches or from the suggestion box, and digitised and anonymised them. Suggestions were then handed over to floor-wise evaluation committees, which consisted of four to five nominated floor managers from different departments. To ensure an efficient evaluation process of suggestions, each committee received a two hour training in January 2020 about the suggestion process and possible criteria to evaluate suggestions (see curriculum at the bottom of Table 1.C.2). The training also included practical team exercises on evaluation of suggestions. Evaluation committees selected three winning suggestions per month and submitted a nomination form to managers for approval (see nomination form template in Figure 1.C.5). The approved winning suggestions were announced by a floor manager, rewarded with prizes, and displayed with the picture and names of winners and the names of supervisors on each floor's notice board. Pictures in Figure 1.C.6 present the design of the display. The process was "semi-blind" in the sense that the workers' names were only anonymous before winning. The winning suggestions were then to be implemented by the factory.

Simultaneously, an industry expert anonymously evaluated suggestions for research purposes. The expert split the suggestions into those that identified a problem without offering a solution, and those that suggested a solution. Evaluations by the factory's evaluation committees and the industry expert show discrepancies (see display of winning suggestions and corresponding expert assessments

in Table 1.C.4) in this dimension. Suggestions which had been assessed as only identifying a problem/behaviour were in fact selected as winning suggestions. These discrepancies might arise from the evaluation committees not only selecting implementable suggestions, but also suggestions containing (not directly implementable) good thoughts. Finally, suggestions and their evaluations were translated into English.

1.2.3 Timeline

Figure 1.2 shows the project's timeline. Every worker assigned to the lunch intervention was supposed to receive lunch once per month over the course of five months. The initial design had to be shortened due to Covid-19, and workers participated in either three or four lunches. The first lunch took place on 26th December 2019. The majority of workers received three lunches between January and March 2020, with only the first of four sub-groups receiving a fourth lunch. I define the time between a lunch of sub-group one and the day before the subsequent lunch of sub-group one as a "round". In each round, each of the four sub-groups received one lunch. The exact lunch dates and rounds can be found in panels a and b of Table 1.C.1. Absent workers received lunch with the next sub-group one week later. The control group was given a take-away lunch in two cohorts on the 21st and the 26th of January 2020 and on the 3rd and 12th of February in case they were absent in January. The motivational training for supervisors took place on the 6th of January 2020 and the technical training on the 4th of March 2020. The first SMS reminder was sent out on the 13th of January (see panel c of Table 1.C.1), with the first huddle scheduled to take place on the 14th of January. Hence, the second managerial intervention started later than the first managerial intervention. In regressions estimating the effect of the second managerial intervention or of both interventions jointly, I drop all observations prior to the 14th of January 2020.

1.2.4 Randomised assignment

The randomisation for the lunch intervention was carried out at the worker level in pairs of workers. The randomisation for the huddle intervention was carried out at the production line level.

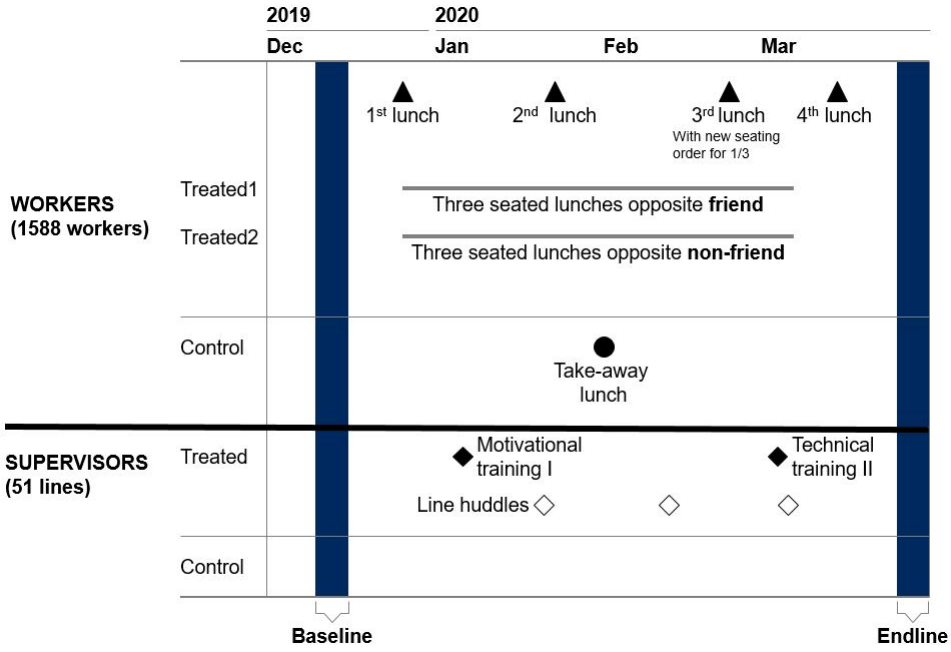


Figure 1.2: Timeline

Notes: Figure 1.2 shows the timeline of the project between December 2019 and March 2020. Baseline surveys were conducted in mid-December 2019, and endline surveys at the end of March 2020.

Randomisation into lunch treatment

The randomisation of 1,599 workers into lunch and non-lunch groups was implemented at the pair level (dyads of workers). Treatment was stratified to have a similar number of friend and non-friend dyads in both groups. Dyads in the control group were never paired in reality. I carried out the randomisation in a six-step approach that I describe briefly here and in detail in the Appendix Section 1.D.

First, I split the sample into workers who, in the baseline survey, name (or are named by) at least one friend (sample one) and those who do not name any other worker with whom they talk (sample two). Second, I paired workers in sample one with a friend and in those sample two with a random worker, a 'non-friend'. To achieve the highest number of possible matches in a scarce network, I first assigned friends to workers who had the lowest number of friends (either those they named or those who named them). Third, I randomised the dyads of friends and non-friends into treatment and control groups. I stratified by sample, production line, and an indicator equal to one if one of the workers per dyad had already been asked extended survey questions. Fourth, I re-balanced treatment assignments

to achieve balance on the maximum experience index per dyad across treatment groups and on the percentage of treatment dyads across lines. The reason to re-balance on the former is that I expect the maximum experience per dyad to drive the quality of suggestions. The experience index reflects a worker's maturity and skillset, and is measured by an Anderson (2008) index of the number of factories worked at, grade, tenure in the factory, tenure in the industry and the educational level. The reason to re-balance on the latter is that I want to avoid effects of the second managerial intervention to be driven by how many workers of the lunch group participate at one huddle. Fifth, I randomly selected 50% of the friend dyads in sample one to be re-paired with a non-friend on the same floor. This step was carried out to be able to compare friend and non-friend dyads within sample one, i.e. within workers who have friends. The final number of workers per treatment assignment is shown in Table 1.D.6. Sixth, after lunch rounds one and two, I randomly selected a third of all pairs assigned to lunch to be seated opposite a new non-friend, as shown in Table 1.D.8. I balanced on the turnover of seating partners. Finally, I randomly choose a seating order for the dyads along the tables in the cafeteria.

Randomisation into huddle treatment

The randomisation into the huddle treatment took place at the line-level. 51 production lines were randomly selected into huddle and non-huddle lines, using the production floor as a stratification variable. The randomisation was carried out 1000 times. For each iteration, I identified the minimum p-value of two-sided t-tests of mean equality between treatment and control lines per balancing variable, namely the number of workers per supervisor, efficiency, alteration rate, and absenteeism. T-tests are conducted separately for production lines on three low- and three high-performing floors, where performance is based on a ranking of all floors with respect to efficiency, alteration rates and absenteeism. This step was undertaken to avoid a treatment assignment of only low-performing lines on some floors and only high-performing lines on other floors. Out of the assignments with a minimum p-value of > 0.20, I chose one at random.

Selection of survey participants

The research team attempted to survey all workers included in the experiment at baseline and endline with a short set of questions. A subset of workers were selected for an extended version of the baseline survey that was conducted in part at the factory and in part on the phone. I randomly selected three operators and one helper per production line, or 204 workers on the 51 production lines, to receive the extended survey. Because of Covid-19, the extended version of the endline survey was conducted entirely on the phone only. I selected at random from each production line two workers from the control group, one worker from the lunch group seated opposite a friend and one worker from the lunch group seated opposite a non-friend. A worker's production line was defined as the line she spent most of her time on between January and March 2020 based on data on line movements provided by the factory.[15] As it was much harder to reach workers at endline, I randomly selected one additional reserve worker per initially selected worker on the same production line and with the same treatment assignment. Enumerators were asked to call reserve workers only if the initially selected worker was unable to be reached after nine attempts at different points in time.[16] All supervisors on the six production floors participating at the experiment were surveyed.

1.3 Data collection and measurement

1.3.1 Data sources

The analysis draws upon four main sources of data: Intervention output, administrative data, survey data and focus groups. First, I obtained the intervention output, i.e. submitted suggestion papers filled out by workers, huddle forms filled out by supervisors and nomination forms filled out by the evaluation committees. Suggestion papers were collected by the research team every week at the end of each lunch session, and also from the suggestion box located on the production

[15] If there were not enough workers left per group per line due to high turnover, I added reserve workers who were on the respective line at baseline. This was the case for two lines for the lunch group of workers seated opposite friends, and for three lines for the lunch group of workers seated opposite non-friends.

[16] In the control group, 68% were surveyed from the initial list, in the friend treatment group 74% and in the non-friend treatment group 76%.

floor. Huddle forms, which give information about the date, the participants and the content of line huddles, had to be filled out by supervisors and handed over to the IE (Industrial Engineering) officer for signed approval. The IE department then delivered the forms to the research team. Nomination forms with information about three nominated winners per floor had to be filled out by the floor-wise evaluation committees and handed over to the management, which then delivered the forms to the research team.

Second, the analysis uses various factory datasets maintained by different departments. The production department provided daily production data including efficiency and quality information. The HR department provided monthly salary data, daily job card data and skills data. The salary data contains information about the workforce, such as the day of joining and the day of exiting the factory. The daily time stamps of each worker with information about the time of entering and of leaving the factory each day is captured in the job card data. The skills matrix data contains information about each process a worker can perform and the time it takes to perform the respective process.[17] The IE department provided line movement data with the daily working location of each worker. Due to incompleteness of the data[18], I draw upon intervention and survey data to complement the line location of workers (see detailed description in Appendix Section 1.B).

Third, the analysis uses survey data of line operators and line supervisors on treatment and control lines collected at two points in time: Shortly before the experiment in early December 2020 (baseline) with the help of eleven enumerators, and shortly after the experiment in mid-March 2020 (endline) with the help of twelve enumerators. The baseline survey consists of one in-person part inside the factory and another phone part. The in-person survey was used to ask about demographics and characteristics, perceptions about the suggestion box and the suggestion culture. The phone survey was used to ask more delicate questions about job satisfaction, feelings and fears. The endline survey was conducted on the phone only, as enumerators were not able to visit the factory due to Covid-19. Both the baseline and the endline survey for line operators consist of a short

[17] The IE team records this data using a stopwatch when a worker joins the factory.

[18] 60% of the days during the intervention period are missing, and 159 out of 1,599 workers are not included in the data at all.

version for all workers, and a long version for randomly selected workers. For supervisors, there was only one version.

Finally, I organised three focus groups with workers, lower/middle managers and higher-level managers each in November 2019 at the end of the pilot experiment. The focus group with workers was comprised of eight workers from the treatment and eight workers from the control group. An IE officer, quality inspector, line supervisor and line chief participated in the focus group with lower/middle managers. The focus group with higher-level managers included four managers, i.e. the General Managers from the Production and the IE department, and the Deputy General Managers from the Quality and the HR department. The purpose of these focus groups was to obtain qualitative, descriptive information on the constraints for workers to submit suggestions and how they felt about the lunch sessions and involvements of supervisors. Although this is a small sample, the focus groups provide richness and a deeper understanding into the mechanisms through which the treatment is expected to have an impact.

1.3.2 Primary and secondary outcome variables

Primary outcomes

I analyse two groups of primary outcome variables which I expect to change in the short-run as a result of the intervention: First, I estimate treatment effects on idea-driven voice, i.e. the submitted suggestions of workers. Second, I estimate treatment effects on worker well-being and firm performance. These two groups of outcomes are important, as both could independently be beneficial for the firm and workers both in the short- and in the long-run. On the one hand, even if suggestions were meaningless, giving workers a voice through encouragement of suggestions might make them feel more listened to and cared for. This empowerment might directly affect worker well-being and/or indirectly affect firm performance through higher motivation and effort of workers in the short-run. On the other hand, even if worker well-being and firm performance was not affected, collecting additional suggestions might be worthwhile in the short-run if their implementations have the potential to improve firm performance in the long-run[19], in line with

[19] Long-run effects are not part of the analysis in this paper.

the Japanese production philosophy *Kaizen*.

i. Idea-driven voice. I analyse treatment effects on three outcomes related to idea-driven voice: Quantity, quality and variety of suggestions. The quantity of suggestions is the total number of suggestions per worker per round, where round is the time in which a treated worker received one lunch. The quality of suggestions is measured with an index score on a scale of one to ten. I use two versions of the quality index to take into account the trade-off between quantity and quality: The first considers only the quality of *submitted* suggestions, with the index starting at one point and workers not submitting any suggestion coded as missing. The second codes workers not submitting any suggestion as zero, so is defined for all workers. In the second version, a bad suggestion is considered a better outcome than no suggestion at all.

Points are based on an expert's anonymous evaluation, in which the identity of the author of the suggestion is not displayed. The ten-point scale is constructed as follows: Each suggestion is classified into one out of six types, where type 1 is an incoherent suggestion, type 2 a general observation/no problem, type 3 a poorly communicated problem, type 4 a well communicated problem, type 5 a poorly communicated idea, and type 6 a well communicated idea. Types 1 to 4 are not further assessed, and receive one to four points, respectively. Types 5 and 6, i.e. ideas, are further evaluated and assigned between one and five additional points. Thus, type 5 can have between five and nine points, and type 6 between six and ten points in total. The additional points for type 5 and 6 are based on the sum of five criteria evaluated on a scale from one to five (implementation level, impact on worker well-being, impact on efficiency, implementation difficulty, and financial resources) and one criterion on a scale from one to two (level of effects). The sum is clustered into five ranges, which make up the additional points. The five ranges are between 1-19, 20, 21-22, 23, and 24-27 points such that each range has a similar number of suggestions.

Furthermore, I differentiate between worker components and firm components which indicate the extent to which a suggestion is beneficial for workers or the firm, respectively. Worker components include the criteria implementation level

and impact on worker well-being, whereas firm components include impact on efficiency, level of effects, implementation difficulty, and financial resources.

The variety - or non-repetition - of suggestions is measured as the number of problems and ideas per round (suggestions of type three to six) which have not been mentioned by the respective worker in any previous round. Variety can only be measured for workers who submitted more than one problem and/or idea. To identify whether suggestions have been mentioned beforehand, suggestions are classified threefold according to different levels of granularity (from least to most granular): Category, subcategory and suggestion. The category here does not necessarily match with the area of improvement a worker could tick on the suggestion paper. A detailed overview of the classifications into category, subcategory and suggestion is shown in Appendix Figure 1.E.8. I report variety based on suggestions, and show the less precise measures of variety based on category or subcategory as robustness checks in the Appendix. As a further robustness check, I generate an additional measure of variety based on latent semantic analysis (LSA), a technique in natural language processing which allows to compare word vectors (Schwarz, 2019).

As an additional outcome, I measure the content of suggestions, which is also based on classifications. I generate an indicator for each category which is equal to one if the suggestion is classified as the respective category. The content of suggestions can obviously only be assessed for submitted suggestions.

ii. Firm performance and worker well-being. I analyse treatment effects on four families of outcomes, related to workforce, production, job satisfaction and mental health. The first two are directly related to firm performance and the second two are directly related to worker well-being. To reduce the number of hypotheses tested, I use a standard 'omnibus' approach, i.e. I group outcomes into families of outcomes, construct an index for each family, and test whether each index is affected by the treatments. For this approach, I first re-code all outcomes such that higher values correspond to a similar direction of change within an index. I then standardise the outcomes to have mean zero and standard deviation one. I finally construct a weighted index of outcomes, using the method of Anderson (2008). I calculate the average of the standardised constituent outcomes, weighted by the

inverse covariance matrix. The family indices are comprised of the following constructed variables (listed in more detail in Table 1.E.9):

- **Workforce index:** Absenteeism, punctuality and working hours
- **Production index:** Efficiency, alterations and reject rates
- **Job satisfaction index:** Job satisfaction and support, turnover intentions and work attitudes
- **Mental health index:** Feelings at work, general anxiety disorder and workplace harassment

Additional secondary outcomes

I also test a wide range of secondary outcomes. Secondary outcomes are used to explain potential mechanisms behind the results. In particular, I look at the level of engagement and support for making suggestions, perceptions about constraints, beliefs about the efficacy of the suggestion box, and perceptions about the workplace culture. Secondary outcomes are based on baseline and endline survey questions. Appendix Figure 1.E.9 displays index components for secondary outcome variables. Indices are also constructed using the method of Anderson (2008).

1.3.3 Empirical specification

To estimate the treatment effects on the primary and secondary outcome variables, I calculate the intent-to-treat (ITT) estimates using four simple regression models. I also calculate local average treatment effects (LATE), shown in the Appendix.

Intent-to-treat estimates

To test the effect of the first and second managerial interventions separately, I use the following ANCOVA specification:

$$Y_{i,l,t} = \alpha + \beta_1 T_{i,l} + \theta Y_{i,l,t=0} + \gamma_{i,l} + \epsilon_{i,l,t} \tag{1.1}$$

where $Y_{i,l,t}$ is the outcome variable of interest for worker i on line l at time t. $T_{i,l}$ is an indicator for random assignment to the treatment, which is either the first

managerial intervention $Lunch_{i,l}$ or the second managerial intervention $Huddle_{i,l}$. $Y_{i,l,t=0}$ is a control for the baseline value of the outcome variable. $\gamma_{i,l}$ is a stratum indicator (experience index), and $\epsilon_{i,l,t}$ is the residual. β is the coefficient of interest.

To test the interaction effects of the lunch and huddle treatment, I estimate the following regression model:

$$Y_{i,l,t} = \alpha + \beta_1 Lunch_{i,l} + \beta_2 Huddle_{i,l} + \beta_3 Lunch_{i,l} * Huddle_{i,l} + \theta Y_{i,l,t=0} + \gamma_{i,l} + \epsilon_{i,l,t} \tag{1.2}$$

where $Lunch_{i,l}$ is an indicator for random assignment to the lunch treatment, $Huddle_{i,l}$ is an indicator for random assignment to the huddle treatment, and $Lunch_{i,l} * Huddle_{i,l}$ is an indicator for random assignment to both the lunch and the huddle treatment. The notation for equation 1.2 is otherwise analogous to that for equation 1.1. In this specification, β_1 is the estimated treatment effect of the lunch treatment, β_2 is the estimated treatment effect of the huddle treatment, and β_3 is the additional effect of receiving both treatments.

To test for heterogeneous treatment effects of each treatment separately, I use the following regression model:

$$Y_{i,l,t} = \alpha + \beta_1 T_{i,l} + \beta_2 HET_{i,l} + \beta_3 T_{i,l} * HET_{i,l} + \theta Y_{i,l,t=0} + \gamma_{i,l} + \epsilon_{i,l,t} \tag{1.3}$$

where $HET_{i,l}$ is an indicator for the above-median baseline value of an interaction variable. The notation for equation 1.3 is otherwise analogous to that for equation 1.2. In this specification, β_1 is the estimated treatment effect on workers with a below-median baseline value of the interaction variable, $\beta_1 + \beta_3$ is the estimated treatment effect on workers with an above-median baseline value of the interaction variable, and β_3 is the estimated difference between these two treatment effects.

I also test the effects on lunch treatment days using the following regression specification:

$$\begin{aligned} Y_{i,l,t} = \beta_0 + \beta_1 td_0 + \beta_2 td_1 + \beta_3 td_2 + \beta_4 td_3 + \beta_5 td_4 + \beta_6 td_5 + \beta_7 td_6 + \beta_8 td_{-1} \\ + \beta_9 td_{-2} + \beta_{10} td_{-3} + \beta_{11} td_{-4} + \beta_{12} td_{-5} + \beta_{13} td_{-6} + \theta Y_{i,l,t=0} + \delta_t + \epsilon_{i,l,t} \end{aligned} \tag{1.4}$$

where td_0 is an indicator for the lunch treatment day. A positive number in the subscript indicates lags (number of days after the treatment day) and a negative number in the subscript indicates leads (number of days before the treatment

day). β_2 to β_7 are the treatment effects on the first day up to the sixth day after the treatment day, whereas β_8 to β_{13} are the treatment effects one to six days before the treatment day.

Standard errors are clustered at the pair level in regression models which test the treatment effects of the lunch treatment only. Standard errors are clustered at the line level in regression models which test treatment effects of the huddle treatment, or treatment effects of the lunch and huddle treatment jointly. All regressions include floor fixed effects to control for any time-invariant differences across floors. Time fixed effects are also included unless the data is collapsed at the individual level. Time fixed effects are in rounds for innovation-related outcomes, in weeks for production-related outcomes, and in days for workforce-related outcomes. Regressions on index variables based on survey data, i.e. job satisfaction- and mental health-related outcomes, as well as all secondary outcomes do not include time fixed effects.

As the lunch treatment started on the 26th of December 2020 and the huddle treatment only on the 14th of January 2020, I restrict the data to all observations after the huddle start date in regression models which test the treatment effects of the huddle treatment, or the treatment effects of the lunch and huddle treatments jointly. In regression models which test the treatment effects of the lunch treatment only, I include all observations.

I report a p-value for each family of outcome variables. I also separately report effects on individual variables. To account for the fact that I am testing multiple hypotheses, I consider more conservative p-values. For each individual variable, i.e. subcomponent of primary outcomes, I report a p-value and False Discovery Rate (FDR) q-values calculated within the family to control for the expected proportion of rejections that are false positives. I do not correct for multiple hypothesis testing when looking at secondary outcomes, as this analysis is informative for exploratory analysis of additional impacts, robustness checks and mechanism analysis, not the main impact.

Local average treatment effects

I also estimate local average treatment effects (LATE) using the randomised assignment to treatment to instrument for actually receiving the treatment. I use random

assignment to the lunch treatment as an instrument for actual participation at at least one lunch, and the random assignment to the huddle treatment as an instrument for actual participation at at least one line huddle. The participation at line huddles is calculated by drawing upon two data sources. First, the exact dates of conducting a huddle per line are identified using information from line huddle templates which supervisors had to submit after each huddle. Second, the exact line position of each worker per day is identified using information from the line movement data. Hence, I infer whether a worker was working on a line on which a line huddle was conducted on a particular day. I then generate a cumulative actual line huddle indicator which is equal to one if a worker has participated at at least one line huddle in the past. Depending on the time structure of the data, I collapse the cumulative actual line huddle indicator by the respective time periods, e.g. by week, month or round. For production data at the line-level, I use the intensity of treatment per line as the actual treatment. It is calculated by dividing the total number of treated workers by the total number of all workers per line and day.

1.3.4 Integrity of the experiment

Baseline Balance and Summary Statistics

I confirm the validity of the randomisation by performing a balance test, results of which are shown in Table 1.1 for primary outcomes and in Table 1.2 for secondary outcomes and characteristics of workers. Only two out of twelve primary outcome variables (absenteeism and general anxiety) in Table 1.1 are different at the 10% level, and none of the secondary outcome variables or characteristics of workers in Table 1.2. A joint orthogonality tests of worker characteristics for each treatment and outcome group shows that overall balance cannot be rejected.

TABLE 1.1. BALANCE TABLE OF PRIMARY OUTCOME VARIABLES

	Control (1)	Lunch (2)	Huddle (3)	p (1=2)	se (Δ)	p (1=3)	se (Δ)	p (1 = 2 ∪ 3)	se (Δ)	N
Workforce-related outcomes										
Workforce index	-0.01	0.02	0.01	0.03	(0.02)	0.73	(0.02)	0.06	(0.02)	1532
Absenteeism	0.02	0.03	0.02	0.03	(0.00)	0.33	(0.00)	0.02	(0.00)	1532
Punctuality	4.89	4.76	5.04	0.26	(0.24)	0.24	(0.29)	0.47	(0.31)	1531
Working hours	11.04	11.03	11.03	0.63	(0.03)	0.60	(0.04)	0.20	(0.04)	1532
Production-related outcomes										
Production index	0.07		-0.01			0.38	(0.10)			51
Efficiency	0.53		0.53			0.61	(0.01)			51
Alteration rate	0.08		0.08			0.58	(0.00)			51
Reject rate	0.00		0.00			0.35	(0.00)			51
Job satisfaction-related outcomes										
Job satisfaction index	0.14	0.09	0.05	0.86	(0.17)	0.65	(0.20)	0.40	(0.24)	204
Job satisfaction	0.34	0.33	0.31	0.89	(0.13)	0.79	(0.16)	0.91	(0.18)	204
Quit intentions	-0.08	-0.17	-0.23	0.81	(0.18)	0.42	(0.21)	0.28	(0.28)	202
Job attitudes	-0.04	-0.00	0.00	0.58	(0.12)	0.50	(0.13)	0.90	(0.16)	202
Job satisfaction supervisors	0.00		-0.22			0.37	(0.25)			94
Mental health-related outcomes										
Mental health index	0.19	0.05	-0.06	0.54	(0.16)	0.11	(0.20)	0.04	(0.21)	202
Feelings at work	0.16	0.07	-0.01	0.77	(0.17)	0.29	(0.22)	0.16	(0.21)	202
General Anxiety	0.11	-0.01	-0.11	0.61	(0.13)	0.04	(0.14)	0.01	(0.15)	202
Workplace harrassment	0.02	-0.02	0.02	0.90	(0.17)	0.80	(0.20)	0.62	(0.26)	202

Notes: Table 1.1 shows balance checks for primary outcome variables across treatment groups. Columns 1 through 3 show sample means in the control group (1), lunch group (2), and the huddle group (3), respectively. Columns 4 through 9 shows p-values and standard deviations of tests for equality of means between the lunch group, the huddle group, and the lunch and huddle group combined, compared to the control groups. P-values are based on standard errors clustered at the worker level for the lunch group, and at the line-level for the huddle group, and the lunch and huddle group combined. The data period for the balance test is three months prior to the intervention start date.

TABLE 1.2. BALANCE TABLE OF WORKER CHARACTERISTICS AND SECONDARY OUTCOME VARIABLES

	Operators										Supervisors				
	Control (1)	Lunch (2)	Huddle (3)	p (1=2)	se (Δ)	p (1=3)	se (Δ)	p (1 = 2 ∪ 3)	se (Δ)	N	Control (4)	Huddle (5)	p (4=5)	se (Δ)	N
Panel A: Characteristics															
Female share	0.79	0.77	0.79	0.69	(0.06)	0.62	(0.05)	0.40	(0.06)	204	0.24	0.32	0.41	(0.09)	94
Age	25.94	25.82	26.13	0.70	(0.70)	0.59	(0.66)	0.91	(0.76)	204	30.44	29.77	0.58	(1.20)	94
Experience index	0.01	-0.09	-0.01	0.24	(0.14)	0.83	(0.14)	0.80	(0.18)	204					
Nb factories	0.90	0.75	0.98	0.15	(0.20)	0.41	(0.19)	0.81	(0.19)	204	2.12	2.41	0.58	(0.53)	94
Grade	4.87	4.99	4.85	0.29	(0.19)	0.71	(0.19)	0.73	(0.24)	203					
Tenure factory in years	2.81	2.84	2.61	0.89	(0.48)	0.46	(0.48)	0.94	(0.65)	204	5.66	5.28	0.60	(0.74)	94
Tenure industry in years	4.64	4.88	4.72	0.64	(0.63)	1.00	(0.61)	0.60	(0.76)	204	10.07	9.80	0.79	(1.01)	94
Education	8.53	8.65	8.22	0.58	(0.51)	0.27	(0.50)	0.82	(0.64)	204	6.66	9.70	0.10	(1.83)	94
Processes	4.76	4.86	5.16	0.99	(0.54)	0.39	(0.69)	0.37	(0.66)	140					
Days per line	54.79	55.13	53.49	0.72	(3.39)	0.58	(3.16)	0.71	(3.80)	204	549.40	632.48	0.63	(171.83)	74
Lines	3.59	3.41	3.64	0.25	(0.28)	0.69	(0.26)	0.91	(0.34)	204	1.12	1.11	0.92	(0.07)	94
Supervised workers											18.18	19.16	0.49	(1.40)	94
Difference L-C (in %)											0.19	0.15	0.12	(0.03)	94
Panel B: Channels															
Engagement and support	-0.02	0.12	0.05	0.22	(0.16)	0.74	(0.19)	0.29	(0.20)	204	-0.36	-0.11	0.26	(0.23)	94
Perceived constraints	-0.15	-0.20	-0.39	0.87	(0.19)	0.15	(0.23)	0.13	(0.23)	172	0.20	-0.03	0.20	(0.18)	94
Perceived efficacy	0.08	-0.03	0.03	0.30	(0.16)	0.78	(0.20)	0.66	(0.21)	204	-0.17	0.01	0.35	(0.20)	94
Communication	0.26	0.31	0.16	0.53	(0.19)	0.35	(0.19)	0.64	(0.18)	204	0.15	0.37	0.42	(0.27)	94
Workplace culture	-0.08	-0.21	-0.18	0.33	(0.15)	0.60	(0.17)	0.18	(0.19)	204	0.07	-0.14	0.38	(0.23)	94

Notes: Table 1.2 shows balance checks for main indices from survey data from line operators (columns 1 to 10) and line supervisors (columns 11-15) across the treatment groups. Columns 1 through 3 show sample means for line operators in the control group (1), lunch group (2), and the huddle group (3), respectively. Columns 4 and 6 show p-values of tests for equality of means between the lunch and the huddle groups compared to the control groups, respectively. Column 8 shows the corresponding p-values for comparisons between the control group and the lunch and huddle group combined. Columns 11 and 12 show sample means for line supervisors in the control group (11) and the huddle group (12). Column 13 shows p-values of tests for equality of means between the control and the huddle groups. P-values are based on standard errors clustered at the worker level for the lunch group, and at the line-level for the huddle group and the lunch and huddle group combined. The test for joint significance of differences between the control and lunch group is not rejected (p-value of the F-statistic of the omnibus balance test: 0.29), similarly for differences between the control and huddle group (p-value of the F-statistic of the omnibus balance test: 0.58) and between supervisors in the control and huddle group (p-value of the F-statistic of the omnibus balance test: 0.11)

A few characteristics are worth highlighting (see Table 1.2): Around 80 % of the sample workers are female with an average age of 26 years, whereas only 28% of the supervisors are female with an average age of 30 years. Workers have completed on average 8 years of schooling which implies that they should be able to read and write. In Bangladesh, students can go to a senior school and then to a high school after nine years of schooling. Less than 2.5% of the sample has completed only two years of schooling and hence, might not be able to write suggestions themselves. For these workers, external facilitators or supervisors could be asked to write down suggestions. Supervisors have also completed on average 8 years of schooling which is not surprising given that most supervisors are recruited from a pool of experienced operators. Workers are able to perform on average 5 different sewing processes, and work on average on 3.5 different lines over a 3-months period.

Compliance

Non-compliance is only one-sided for the lunch treatment and not very strong. Control workers never receive any lunch, and less than ten percent of workers assigned to the lunch treatment did not receive any lunch. Therefore, the assigned lunch and actual lunch variables are almost identical, and LATE and ITT effects are very similar.

On the contrary, non-compliance for the huddle treatment is two-sided and stronger due to two sources: Supervisors incorrectly conducting or not conducting line huddles, and workers extensively moving across lines. Regarding the former source, an evaluation of the line huddle forms indicates that non-compliance is low (see Table 1.C.3). 92% of the treatment lines (23/25) correctly conducted at least one huddle in January 2020. One treatment line conducted its first huddle mid-February, and one failed to conduct any huddle in the entire treatment period. 15% of the control lines (4/26) incorrectly conducted a huddle by mid-February 2020, indicating some minor non-compliance on control lines. The larger source of non-compliance comes from the fact that workers moved across lines. Figure 1.F.10 shows the distribution of the number of lines per worker during the intervention period and Figure 1.F.11 shows the distribution of the number of days working on

the same line. Workers were assigned to an average of 3.4 lines during the huddle treatment window, and the average maximum number of days per worker per line during the 90-day period between January and March 2020 is 56 days. I cannot rule out that workers' line movements are endogeneously related to unobservables. For example, workers might move more often across lines if they can perform multiple processes. These workers might also submit more and better suggestions, implying that workers' line movements are not exogeneous to treatment. In fact, a simple OLS regression in Table 1.3 shows that line movements are increasing (column 1) and the maximum number of days per line are decreasing (column 2) in the number of processes a worker can perform with p-values of 0.11 and 0.085, respectively. Thus, I rely on ITT effects when interpreting results. LATE (IV) effects on primary outcomes are shown in the Appendix .

TABLE 1.3. EFFECT OF SKILL LEVEL ON LINE MOVEMENTS

	Total number of lines	**Max. days per line**
	(1)	(2)
Number of processes per worker	0.035	-0.73*
	(0.022)	(0.42)
Observations	1177	973
R^2	0.26	0.06
Control Mean	3.63	57.32
Floor FE	YES	YES

Notes: Table 1.3 shows an OLS regression of the number of processes each worker can perform (based on skills matrix data) on the total number of lines in column 1 and on the maximum number of days per line in column 2, each based on the intervention period from January to March 2020. All regressions include factory fixed effects. Robust standard errors are in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Attrition

In general, it is much easier to reach workers by conducting in-person surveys in the factory than phone surveys. The baseline survey could be conducted in the factory. Only a small subset of workers was surveyed additional sensitive questions on the phone. Due to Covid-19, the endline survey had to be conducted on the phone only. The survey team made a great effort in following up with as

TABLE 1.4. ATTRITION

	Dep. variable: Attrition		
	(1)	(2)	(3)
Lunch	-0.007		0.001
	(0.752)		(0.969)
Huddle		0.043	0.052
		(0.109)	(0.110)
LunchXHuddle			-0.019
			(0.655)
Observations	1587	1598	1587
R^2	0.00	0.00	0.00
Control Mean	0.27	0.25	0.26
Clusters	794	51	51
p (L=H)			0.12
p (L=LxH)			0.76
p (H=LxH)			0.28

Notes: Table 1.4 shows a linear regression of the treatment indicators on a variable equal to one if the worker was not surveyed at endline. Standard errors are in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

many workers as possible. As the factory was already closed due to the pandemic, and most of the workers went home to their families, it was a lot harder to reach them. The survey team tried to reach each worker up to nine times. In total, 73% of the sample could be re-surveyed at endline. Of concern is whether treatment is correlated with attrition. I test for this in Table 1.4 by regressing a dummy variable indicating if the worker was not found at endline on treatment indicators. I find no significant differences in attrition rates across treatment arms. Above all, most of the analysis using survey data relies on extended survey questions only. As randomly selected workers were surveyed at endline (not necessarily the same as at baseline), there is no differential attrition between baseline and endline. For supervisors, I received an updated list with phone numbers of all supervisors on treated and control lines. I ensured that at least one supervisor was surveyed per line independent of whether he had been surveyed at baseline or not.

1.4 Main outcomes

1.4.1 Impact on idea-driven voice

I begin by examining the impact on idea-driven voice on the quantity, quality, and variety of suggestions. The results are shown in Table 1.5. I first describe the effects of the 'lunch intervention', i.e. structured peer-to-peer lunches, and then analyse additional effects of the 'huddle intervention', i.e. the supervisor training and line huddles.

TABLE 1.5. EFFECTS ON PRIMARY OUTCOME VARIABLES RELATED TO INNOVATION

	Quantity			**Quality**			**Quality (if submitted)**			**Variety**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Lunch	0.56***		0.46***	2.39***		2.04***	-1.17**		-1.28*	
	(0.017)		(0.024)	(0.077)		(0.13)	(0.46)		(0.70)	
Huddle		-0.019	-0.022*		-0.078	-0.12		0.013	0.15	0.068
		(0.019)	(0.011)		(0.11)	(0.076)		(0.17)	(1.09)	(0.052)
LunchXHuddle			-0.011			0.010			-0.11	
			(0.031)			(0.16)			(1.07)	
Observations	6352	4956	4956	6710	5179	5179	1857	1294	1294	543
R^2	0.27	0.06	0.23	0.27	0.06	0.23	0.06	0.07	0.08	0.05
Control Mean	0.03	0.36	0.26	0.09	1.28	0.91	6.10	5.00	5.02	1.11
Round FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Floor FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
p (L=H)			0.00			0.00			0.08	
p (L=LxH)			0.00			0.00			0.48	
p (H=LxH)			0.76			0.49			0.90	

Notes: In columns 1, 4 and 7 standard errors are clustered at the pair level. In all other columns, standard errors are clustered at the line level, and observations prior to the first huddle (14th of January 2020) are excluded. Columns 7 to 10 only include submitted suggestions. Column 10 further restricts observations to individuals who submitted more than once (which are only some of the lunch workers). The quality index (column 4 to 9) consists of six components, out of which two are criteria to evaluate the usefulness of submitted ideas for workers, namely impact on worker well-being and implementation level, and four are criteria to evaluate the usefulness of ideas for the firm, namely financial resources, impact on efficiency, implementation time and risk, and level of effects (separate regressions shown in Table 1.G.14). * p<0.1, ** p<0.05, *** p<0.01

I find that the lunch intervention significantly increases the quantity and relevance of suggestions that are implementable in the firm. Column 1 shows that workers who were not assigned to lunch have a 3% probability of submitting suggestions, and this leaps to 59% for those in the lunch group.

Figure 1.3 illustrates this result by showing the large difference in the number

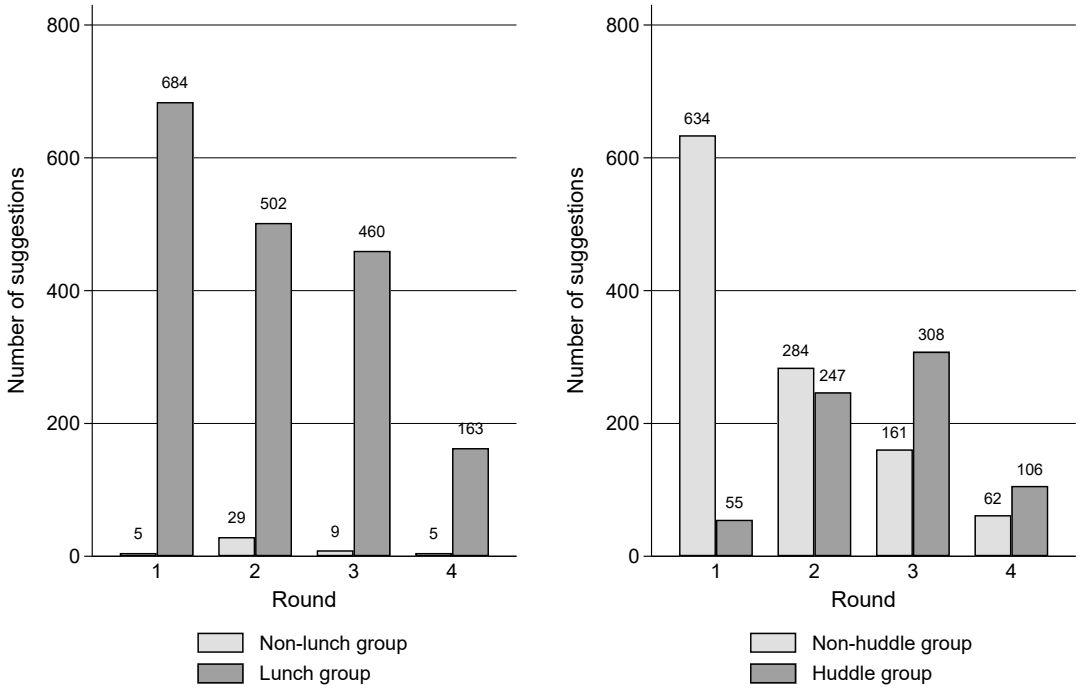


Figure 1.3: Total number of suggestions per round

Notes: The left graph of Figure 1.3 shows the total number of suggestions submitted each round by workers assigned to the lunch treatment and control, whereas the right graph shows the same for the huddle treatment and control workers. If a submitted suggestion paper contains several different suggestions, the paper is split by suggestion. In round four, only a quarter of the lunch group was treated (due to Covid-19), leading to a sharp decrease in submissions. Please note that the number of rounds does not always correspond to the actual number of lunches in case of non-compliance, e.g. if a worker was ill/absent, and that it never corresponds to the number of huddles. Few workers received a huddle before the end of Round 1. Overall, the total number of suggestions from the lunch group over the course of the intervention period is 1,809 as compared to 48 from the non-lunch group. The total number of suggestions from the huddle group after the 14th of January, which is the date by which the first huddle took place (at the end of round one), is 716 across all rounds as compared to 507 from the non-huddle group.

of submitted suggestions per round. The total number of suggestions from the lunch group over the course of the intervention period is 1,809 as compared to 48 from the non-lunch group.

Opposite effects can be seen in respect of the quality of suggestions submitted: The quality is on average 1.2 points lower in the lunch group compared to the non-lunch group as indicated by the significantly negative coefficient in column 7, which is a reduction in quality by 20% compared to the mean of the non-lunch group. Figure 1.4 illustrates this result by showing the distribution of the quality index of suggestions with a larger tail at the top end for the non-lunch group, stemming from a higher quality related to firm components as shown in Table 1.G.14. Overall, this result suggests that there is a positive self-selection of high-quality submissions from workers not assigned to the lunch treatment. It is more costly for them to submit suggestions in terms of required time and effort, so the cost acts as a filter.

Section 1.E described a second version of the quality index, in which suggestions of low quality are considered as better than no suggestion at all (i.e. non-submissions are coded as zero instead of missing). By this measure, the quality index of the lunch group is 24% higher than of the non-lunch group. Figure 1.F.14 illustrates the difference between the two measures of quality, with higher quality of the non-lunch group in the upper right graph and higher quality of the lunch group in the upper left graph. A detailed evaluation of those two measures of quality is described in Subsection 1.4.3.

I also find that workers in the lunch group submit suggestions of much larger breadth, covering diverse topics. In fact, the non-lunch group mainly submits suggestions about leave practices. Some categories, such as work environment/facilities, salary, motion/waiting time, efficiency, or health and safety, are not covered by the non-lunch group at all. The mean of submissions of the non-lunch group and treatment effects for the lunch group by category are shown in the upper graph of Figure 1.5 (also see Appendix Table 1.G.16). As workers in the non-lunch group never submitted more than one suggestion, the variety of suggestions cannot be compared between the lunch and non-lunch group.

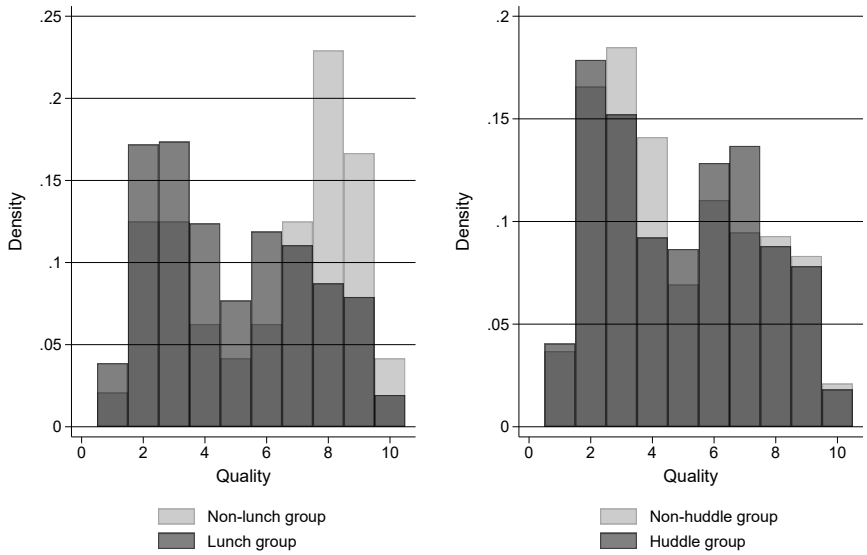


Figure 1.4: Distribution of quality of submitted suggestions

Notes: Figure 1.4 shows the distribution of the quality of submitted suggestions by lunch treatment and control group on the left side and by huddle treatment and control group on the right side. Suggestions classified as incoherent receive a quality index rating of 1, general observations/no problems 2, a poorly communicated problem 3, a well communicated problem 4, a poorly communicated idea 5, a well-communicated idea 6. Ideas are then further evaluated based on five criteria on a scale from one to five (implementation level, impact on worker well-being, impact on efficiency, implementation difficulty, and financial resources), and one criterion on a scale from one to two (level of effects). The sum is clustered into five ranges (1-19, 20, 21-22, 23, and 24-27 points). Points are clustered such that each of the five ranges has a similar number of observations. Those ranges are added to the quality index for poorly or well communicated ideas, such that they range from five to nine or six to ten points, respectively.

The effects of the huddle intervention are less promising. Neither the quantity, quality nor variety of suggestions changes (see columns 2, 5, 8 and 10 of Table 1.5). Results on variety are robust to different specifications of variety based on categories and subcategories, or a latent semantic analysis approach as shown in the Appendix Table 1.G.15. There are also no interaction effects between line huddles and lunches (see columns 3, 6 and 9 of Table 1.5). The negative effect on quantity in column 3 is even marginally significant. However, I find more robust evidence that the huddle treatment influences the content of suggestions. The huddle and non-huddle groups submit a similar number of suggestions, but the huddle group submits significantly fewer suggestions about supervisor skills as shown in the lower graph of Figure 1.5 (also see column 1 of Appendix Table 1.G.16 panel

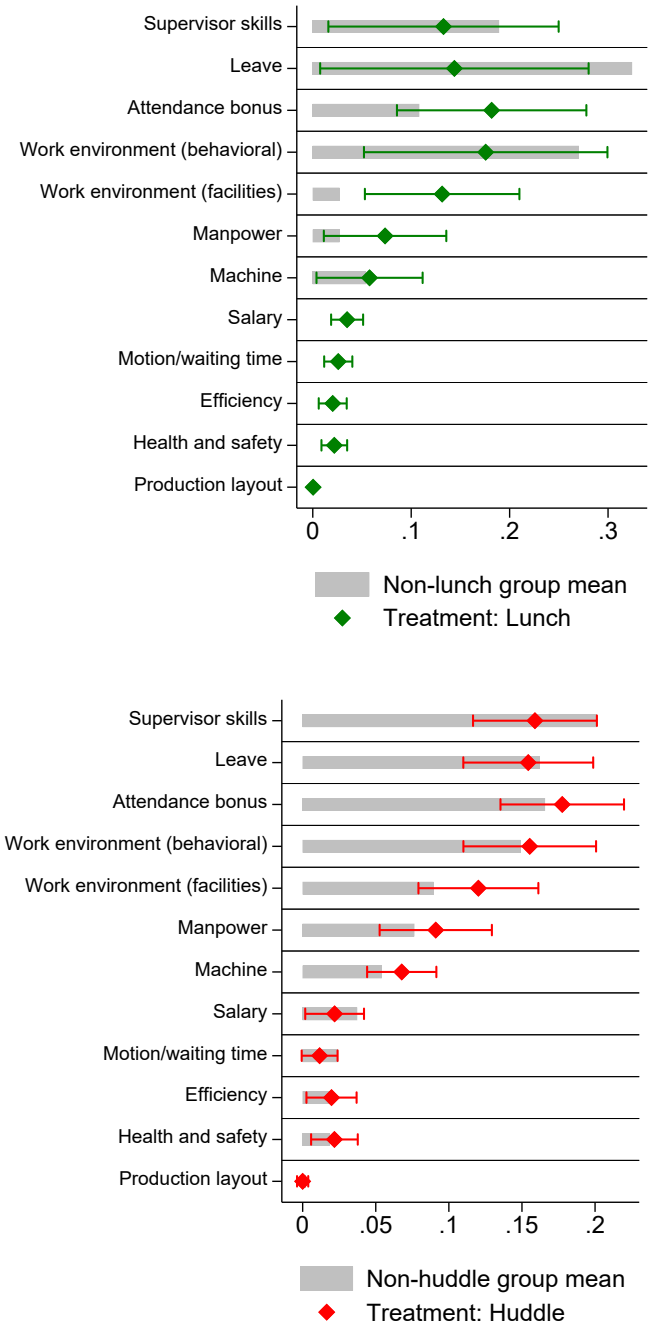


Figure 1.5: Effects on content of suggestions by treatment groups

Notes: Figure 1.5 shows the ITT effects of twelve individual regressions of the treatment indicator on the content variable. The content variable is an indicator equal to one if the submitted suggestions is classified as the respective category. Suggestions are only classified if they are problems or ideas (type three or six). Regressions in the lower graph exclude all suggestions prior to the huddle start date, i.e. 14th of January 2020. 95% confidence intervals are displayed.

B). The category supervisor skills involves suggestions about three subcategories: Skill assignments (e.g. managing work allocations), supervisor gender (e.g. avoiding female supervisors), and skills (e.g. no scolding). Overall, these results suggest that the encouragement of supervisors does not serve as an additional accelerator to make peer-to-peer lunches more efficient, and even undermines workers' critical voice about supervisors' skills.

1.4.2 Impact on worker well-being and firm performance

The impact on firm performance, represented by the workforce index and the production index, and worker well-being, represented by the job satisfaction and the mental health index, is shown in Table 1.6.

TABLE 1.6. EFFECTS ON PRIMARY OUTCOME VARIABLES RELATED TO FIRM PERFORMANCE AND WORKER WELL-BEING

	Firm performance				**Worker well-being**					
	Workforce index			**Production index**	**Job satisfaction index**			**Mental health index**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Lunch	-0.006		-0.003		0.197**		0.305*	0.022		0.099
	(0.720)		(0.909)		(0.036)		(0.087)	(0.855)		(0.469)
Huddle		0.012	0.024	-0.169		0.107	0.221		0.151	0.249
		(0.539)	(0.366)	(0.184)		(0.217)	(0.123)		(0.347)	(0.192)
LunchXHuddle			-0.024				-0.244			-0.190
			(0.528)				(0.235)			(0.383)
Observations	119588	91580	91580	459	195	195	195	195	195	195
R^2	0.09	0.09	0.09	0.47	0.20	0.18	0.21	0.10	0.10	0.11
Control Mean	-0.14	-0.10	-0.10	0.15	-0.23	-0.17	-0.14	-0.13	-0.17	-0.12
Clusters	792	785	785	51	182	51	51	182	51	51
p (L=H)			0.32				0.53			0.43
p (L=LxH)			0.72				0.13			0.38
p (H=LxH)			0.42				0.16			0.22

Notes: Standard errors are clustered at the pair level in columns 1, 5, and 8, and at the line level otherwise. Time fixed effects are included in regressions on the workforce index (in days) and on the production index (in weeks). Huddle exposure is truncated at one (at least one huddle). The cumulative exposure per worker is used in columns 1, 2 and 3, the average cumulative exposure per line in column 4, and the maximum exposure per worker in columns 5 to 10. Indices are constructed using the Anderson (2008) method. Standard errors are in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

I find no short-run effect of the lunch intervention on firm performance, but I find significantly positive effects on worker well-being, in particular on the job satisfaction index with an estimate of 0.2 in column 5 and of 0.3 in column 7 once both treatments are included. A closer look at the components of the job satisfaction index in Table 1.G.13 shows that the positive effect is driven by significantly

improved job attitudes in column 10. This result survives multiple testing correction with a q-value of 0.055. I calculate q-values, the False discovery rate (FDR) adjusted p-values, following the method of Benjamini et al. (2006). The job attitudes sub-index includes questions about the extent to which workers do the best job they can and workers' sense of punctuality. Punctuality is a particularly important job attitude given the sequential nature of the production process, as workers might not be able to start working if another worker on the line has not performed the preceding sewing step. Coefficients on the job satisfaction sub-index in column 8 and turnover intentions sub-index in column 9 are positive, but not significant. The workplace harassment sub-index, a component of the mental health index, also has a significantly positive effect of 0.2. This index includes questions about the frequency of abusive language or uncomfortable situations at work. One must bear in mind that the endline survey was conducted in the midst of the outbreak of Covid-19, which explains why the control mean of the job satisfaction and mental health index is negative at endline (see Figure 1.F.12). The lunch intervention resulted in a much less severe deterioration in the treatment than in the control group. This makes the result of improved worker well-being for the lunch treatment even more promising, as positive effects on other sub-indices such as general anxiety plausibly might have been stronger in the absence of Covid-19. For example, the variance of general anxiety in the control group increases by 79% between baseline and endline, suggesting that statistical power might be lower in this measure.

I also examine effects of the lunch treatment at the daily level by including leads and lags of the treatment day in the regression specification. These effects can only be estimated for variables available at the daily level, i.e. the workforce index components. I find that workers are significantly less absent on days at which they are assigned to receive a free lunch and on the day after (see plot of coefficients in Figure 1.6). This result illustrates that workers make sure not to be absent in order to receive a free lunch gift from the firm and to participate at the suggestion process, which could make them feel valued and cared for. There are no effects on punctuality or working time at lunch treatment days.

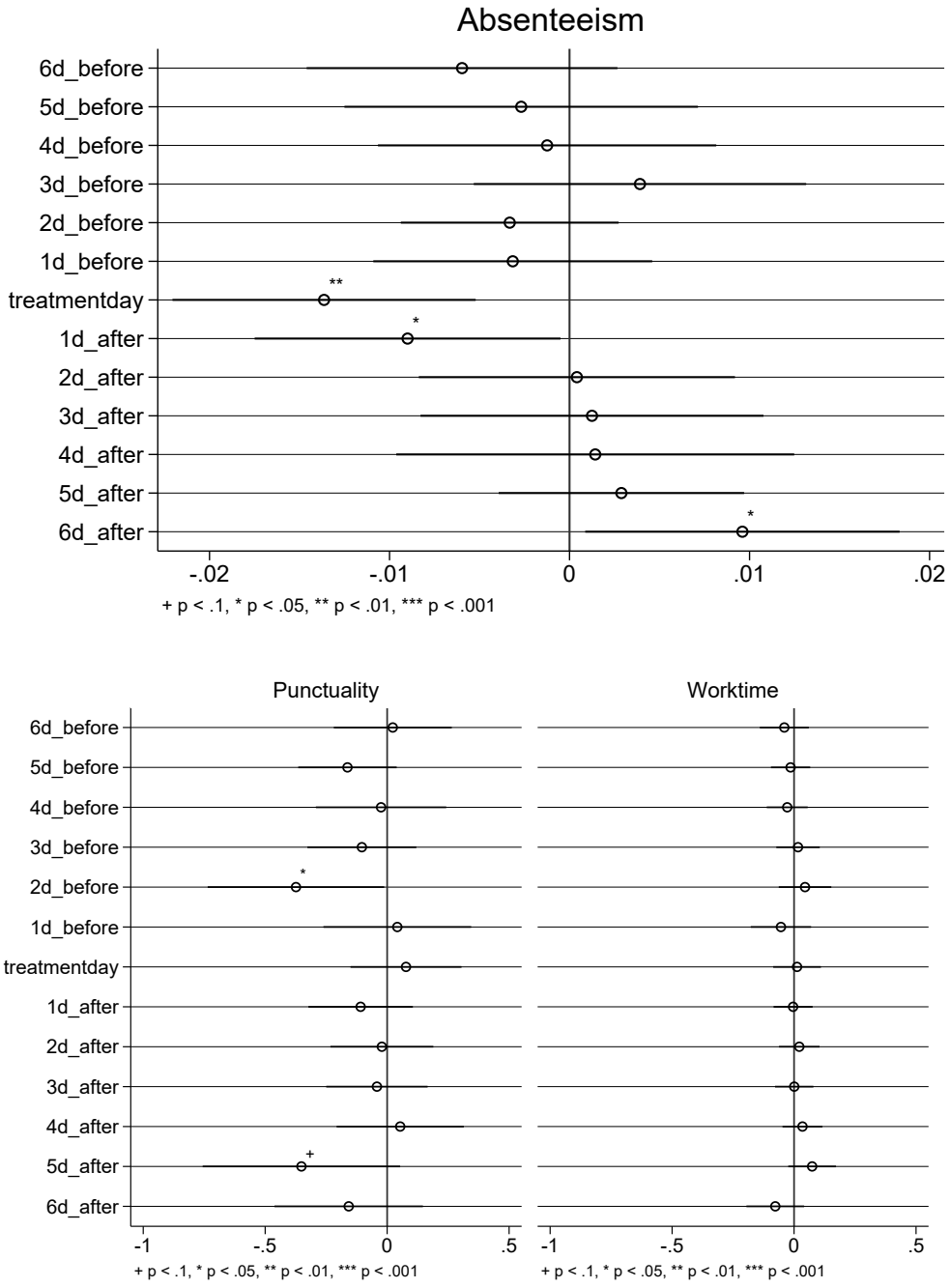


Figure 1.6: Daily effects on workforce index components

Notes: Table 1.6 shows the plot of coefficients of regressions of the lunch treatment days including six leads and lags on primary outcome variables at the daily level. "treatmentday" is an indicator equal to one on assigned treatment days. "Xd_before" is an indicator equal to one for the Xth day before assigned treatment days, and "Xd_after" for the Xth day after assigned treatment days. Absenteeism is an indicator equal to one if the worker was absent on the respective day. Punctuality measures the time a worker arrives before the assigned shift in minutes. Worktime measures the daily working time in hours.

The huddle treatment has no significant effects on any of the four primary outcome variables. Looking at the sub-components in Table 1.G.13, only two of 13 measures are significant, both at the .10 level. Working time increases and alterations decrease, both marginally significantly. Moreover, the coefficients on both working time and alterations are negligibly small and results do not survive multiple testing correcting.

1.4.3 Evaluation of quantity-quality trade-off

In evaluating the lunch intervention, the firm is confronted with a quantity-quality trade-off to either get a higher quantity of suggestions with a larger variance in quality from the lunch group or a lower quantity of suggestions with a higher quality from the non-lunch group.[20] I provide three explanations to conclude that the latter is more beneficial for the firm. First, the lunch treatment makes workers significantly better off in terms of job satisfaction (as outlined in Subsection 1.4.2), which could eventually lead to reduced turnover in the long-run. The highest turnover is usually around the Muslim holiday Eid al-Adha[21], which could not be measured as part of this project due to Covid-19. The lower absenteeism on assigned lunch treatment days is suggestive of workers enjoying to be involved in the suggestion process and to receive a free lunch gift from the firm. Hence, it may even be a better outcome to receive bad suggestions than no suggestions at all. Second, the lunch treatment is a relatively cheap intervention to improve worker well-being. Not only the cost of evaluating suggestions is generally low[22], but also the cost of evaluating suggestions 'in vain' is even lower in the lunch group, as

[20]This trade-off is captured by using two versions of the quality index as described in Section 1.3.2. The first version codes non-submissions as missing, suggesting a preference of quality over quantity. The second version codes non-submissions as zero, suggesting a preference of quantity over quality while simultaneously considering quality, i.e. suggestions of low quality are considered as better than no suggestion at all, but suggestions of high quality are also considered as better than low-quality suggestions.

[21]There are two key Eid's (Celebration Festivals) in Islam: Eid-ul-Fitr, which signifies the completion of the Holy Month of Ramadan; and Eid-ul-Adha, the greater Eid, which follows the completion of the annual Hajj pilgrimage, at the time of Qurbani (sacrifice) (BBC, 2019). They are celebrated around May and August, respectively, although the larger turnover is usually after the second Eid in August.

[22]It takes approximately 30 seconds to assess the type of a suggestions and additional 60 seconds to further evaluate ideas. Assuming that each worker per floor submits one suggestion, i.e. 250 suggestions, and 30% of those are ideas, the floor-wise evaluation committee needs to invest only about 3.3 hours per month.

the likelihood of getting the best suggestions is higher. In fact, 88% of the suggestions judged to be best (blind to treatment status) were from the lunch group.[23] Third, the lunch treatment enables the firm to get a feeling for what workers think and care about through the larger breadth of submitted topics (as outlined in Subsection 1.4.1). Implementing suggestions on those topics could eventually lead to productivity improvements in the long-run. In fact, some workers already noticed the implementation of a few suggestions at endline, such as the instalment of new fans or the repair of broken floor tiles.

On the contrary, there is no evidence that the huddle treatment is beneficial. It has neither effects on firm performance and worker well-being, nor effects on the quantity, quality and variety of suggestions. It reduced the number of suggestions submitted on supervisors' skills and behaviours. The main results hold using LATE (IV) specifications as shown in the Appendix Table 1.G.11 and 1.G.12.[24]

1.5 Mechanisms

In this section, I look at potential mechanisms that can explain results on primary outcome variables. I first look at mechanisms driving results at the extensive margin, i.e. why lunch workers submit at all. I aim to explain the significantly higher number of submissions in the lunch group compared to the non-lunch group. Second, I analyse mechanisms driving results at the intensive margin, i.e. why certain workers submit even more and better suggestions. I hereby focus on the lunch group, as the non-lunch group barely submitted any suggestions at all. Third, I look at mechanisms driving the lower number of suggestions on supervisor skills in the huddle group compared to the non-huddle group. Finally, I analyse potential drivers of job satisfaction.

1.5.1 Why do workers submit suggestions at all?

In this subsection, I analyse possible channels explaining why the lunch treatment induces workers to exert their voice at the extensive margin. First, by design of the

[23]The factory awarded five suggestions from the non-lunch group and 38 suggestions from the lunch group with prizes.

[24]Results are also robust to excluding randomisation sample two (see Section 1.D for explanations of the sample composition).

lunch experiment, time is set aside for workers to jointly discuss suggestions under the help of external facilitators. I therefore expect to find changes in the workers' perceived costs of effort (including opportunity costs of time). As presented in a short framework in the theoretical appendix in Section 1.A, the utility for workers from idea generation is increasing with decreasing costs of effort. I measure perceived costs of effort through workers' perceived ease of generating ideas and ease of submitting ideas.[25] Second, in line with the theoretical framework, I expect effects to be also driven by workers' perceived probabilities of winning, which I measure through an aspirations index.

Table 1.7 presents the results. Related to the perceived costs of effort, I find changes in the ease of submission index, which captures significantly positive effects on both the extent to which workers feel incentivised and comfortable to submit suggestions. I do not find any evidence of changes in the ease of idea generation index. Workers do not find it easier to think about problems and solutions, suggesting that the lunch treatment does not seem to have improved workers' perceived intellectual abilities to generate suggestions.

Related to workers' beliefs, I find that workers in the lunch group have developed significantly greater aspirations to submit suggestions and to win a prize. The survey asked workers to predict the number of suggestions they think to submit next month, and to estimate the likelihood that they will win and have their own idea implemented (detailed overview of all variables in Appendix Figure 1.E.9). The coefficients from regressing the lunch treatment indicator on those survey variables are significant, as shown in table 1.7. For example, the coefficient on the belief about winning and the belief about the implementation a worker's own idea are significant at the 1% significance level and large in size (3.8 and 2.6 standard deviations, respectively). Overall, they are much more confident of their own contribution to the suggestion system. This is also what external facilitators experienced during lunch sessions. Initially, many workers found it indeed challenging to come up with suggestions and were concerned about not being able to say anything. Only after external facilitators constantly and clearly explained

[25]Opportunity costs of time could not directly be measured, but workers themselves did not recognize time as a constraint for idea generation (see "constraints operator" index in Appendix Table 1.H.18).

TABLE 1.7. EFFECTS ON WORKERS' PERCEIVED EASE OF IDEA GENERATION, EASE OF SUBMISSION AND ASPIRATIONS

	Control Mean	Lunch	N
Ease of idea generation index	-0.04	0.01	200
		(0.13)	
Easiness to think about problems	-0.05	-0.01	200
		(0.13)	
Easiness to think about solutions	-0.03	0.02	200
		(0.13)	
Ease of submission index	0.30	0.19***	1164
		(0.06)	
Feeling incentivized to submit idea	0.26	0.17***	1164
		(0.06)	
Feeling comfortable to submit idea	0.28	0.17***	1164
		(0.06)	
Aspirations index	-0.04	0.34***	200
		(0.12)	
Predicted number of own submissions	0.03	0.05*	200
		(0.03)	
Belief about winning	-0.04	0.38***	200
		(0.10)	
Belief about implementation of winner's idea	-0.05	0.22	200
		(0.15)	
Belief about implementation of own idea	-0.09	0.39***	200
		(0.15)	

Notes: Table 1.7 shows ITT effects of the first managerial intervention on workers' perceived abilities, ease of submission and aspirations. Indices are constructed based on Anderson (2008) and include indented standardised variables beneath, which are coded such that higher values correspond to treatment effects in the same direction. Each row is a separate regression. Column 1 shows the mean of workers not assigned to a lunch. Column 2 shows the magnitude of the coefficient with the standard deviation below in parentheses. Column 3 indicates the number of observations N. It is higher for the ease of submission index as corresponding survey questions were asked to all workers at endline. Floor fixed effects are included in the regressions and standard errors are clustered at the pair level. * p<0.1, ** p<0.05, *** p<0.01

the task and intensively motivated workers, workers developed an awareness that they actually have something to say.[26]

I examine a number of different alternative hypotheses to explain the impact of the lunch treatment, but I do not find any compelling evidence for any of these potential explanations. By design, I rule out gift-exchange as a core mechanism (Gneezy and List, 2006; Dellavigna et al., 2018, 2016), as the control group also

[26]External facilitators were explicitly instructed not to provide workers with content for suggestions to rule out that changes in aspirations are driven by technical help.

received an individual take-away lunch with a suggestion paper and a short instruction. I also rule out more effort due to the fact that operators work towards a mission they care for (Dellavigna et al., 2018; Tonin and Vlassopoulos, 2010, 2015). For this mechanism to be at work, the lunch intervention would have to change workers' perceptions about how seriously the firm takes the suggestion box. I argue that this is not the case. Firstly, if it had been the case, I would have expected to see spillover effects on control workers with friends who had lunch. Yet those workers submit a similar number of suggestions as control workers with friends who had no lunch. Secondly, even after the first announcement of awards, which also signals the seriousness of the suggestion box, control workers' behaviour did not change.

Overall, I can conclude that the lunch intervention has successfully addressed impediments to idea generation by changing workers' perceived cost of effort, in the form of ease of submissions, and by simultaneously changing workers' aspirations to participate at the suggestion process. For firms in low-income countries, addressing workers' aspirations seems to be particularly important in unlocking bottom-up information flows across hierarchies, and offers large potential for further research.

1.5.2 Who submits more and better suggestions?

In this subsection, I explore why some workers submit more and better suggestions than others, i.e. at the intensive margin. I mainly focus on the lunch group when analysing results, as the non-lunch group barely submitted any suggestions at all. To examine the determinants of the quantity and quality of suggestions, I look at three areas potentially driving results: Individual dispositions of workers and their peers, individual dispositions of supervisors, and learning over time.

Effects by individual disposition of workers and their peers

In this subsection, I unpack heterogeneous treatment effects by individual characteristics of workers and their peers. I first examine the importance of workers' own dispositions in determining the quantity and quality of suggestions, using the whole sample. In particular, I analyse heterogeneous effects by experience and

gender. I then examine heterogeneous treatment effects by the disposition of seating partners and how those effects differ depending on workers' own disposition, focusing on workers in the lunch group only. In particular, I analyse heterogeneous effects by experience, friendship and gender compositions of seating partners.

Effects by workers' own disposition. I first look at heterogeneous effects by workers' baseline experience index, which reflects a worker's maturity and skillset. This is measured by an Anderson (2008) index of the number of factories worked at, grade, tenure in the factory, tenure in the industry and the educational level, and was used for stratification (see Section 1.2.4 on the randomisation of the lunch treatment). Table 1.8 shows that the lunch treatment is effective for workers with a below-median experience index, but even more effective for workers with an above-median experience index as indicated by the significantly positive coefficients on the interaction term "Lunch * High-exp worker" in columns 1 and 4. Within the lunch group, above-median experienced workers submit on average 0.18 suggestions more per round of 7% higher quality than below-median experienced workers. There are no heterogeneous effects by experience for the huddle treatment (see columns 2, 3, 5 and 6).

Are workers with an above-median experience index more comfortable and confident to state what they think, e.g. because they may fear less the consequences of bad suggestions, or do they possess better abilities[27] to develop suggestions, e.g. because they may have more insights as to what could be changed? I argue that abilities drive the result. As we have seen in Section 1.5.1, the lunch treatment has significantly improved the ease of submission, i.e. the comfort and confidence, but not the ease of idea generation, i.e. the abilities. Thus, I can expect that it takes a high level of confidence for the non-lunch group to actually submit a suggestion. If confidence was the main driver, I would have expected above-median experienced workers to submit more and better suggestions even without being assigned to the lunch group, which is not the case (coefficients on "High-exp worker" in column 1 and 4 are not significant).

To provide further evidence for workers' abilities to drive the results, I explore heterogeneous effects by the number of processes a worker can perform, which

[27] Abilities may come from both tenure and/or intelligence.

TABLE 1.8. HETEROGENEOUS TREATMENT EFFECTS BY WORKERS' EXPERIENCE

	Quantity			**Quality**		
	(1)	(2)	(3)	(4)	(5)	(6)
High-exp worker	-0.0055	0.10***	-0.0040	0.00036	0.47***	0.012
	(0.0069)	(0.027)	(0.012)	(0.028)	(0.12)	(0.044)
Lunch	0.46***		0.37***	1.75***		1.47***
	(0.023)		(0.028)	(0.089)		(0.11)
Lunch * High-exp worker	0.18***		0.17***	0.71***		0.74***
	(0.032)		(0.043)	(0.13)		(0.18)
Huddle		0.0061	-0.013		0.049	-0.035
		(0.025)	(0.014)		(0.098)	(0.052)
Huddle * High-exp worker		-0.044	-0.0091		-0.19	-0.045
		(0.037)	(0.014)		(0.16)	(0.065)
Lunch * Huddle			-0.0061			-0.0089
			(0.041)			(0.14)
Lunch * Huddle * High-exp worker			-0.0059			-0.035
			(0.051)			(0.22)
Observations	6352	4955	4955	6352	4955	4955
R^2	0.28	0.06	0.24	0.25	0.06	0.22
Control Mean	0.02	0.22	0.03	0.05	0.83	0.09
Round FE	YES	YES	YES	YES	YES	YES
Floor FE	YES	YES	YES	YES	YES	YES
Lunch=Lunch*High-exp worker	0.00		0.00	0.00		0.01
Huddle=Huddle*High-exp worker		0.38	0.87		0.29	0.92

Notes: Table 1.8 shows heterogeneous treatment effects by workers' experience. Experience is measured by an Anderson (2008) index of the number of factories worked at, grade, tenure in the factory, tenure in the industry and the educational level. "High-exp worker" means that the respective worker has an above-median experience index. Columns 2, 3, 5, and 6 exclude all observations prior to the start date of huddles, i.e. the 14th of January 2020. A the bottom, I show p-values from comparing each treatment and treatment interaction against each other. Regressions include round and floor fixed effects. Standard errors are clustered at the pair level in columns 1 and 4, and at the line level in columns 2, 3, 5 and 6. Standard errors are shown in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

likewise reflects a worker's skillset.[28] Although skills matrices are missing for approximately 26% of the sample, Appendix Table 1.H.20 shows that results have very similar patterns as in Table 1.8. This is not surprising given that there is a positive correlation of 0.21 between the number of processes and the experience index. Besides, I can see that the huddle is particularly effective for low-skilled workers, i.e. workers with an above-median number of processes (see column 2).

I finally examine heterogeneous effects by workers' own gender. The sample is composed of 25% male and 75% female workers. It would be misleading to rely on results which only take into account a worker's gender[29], given that only 45% of women have an above-median experience index, whereas 65% of men. Hence, I look at heterogeneous treatment effects on the quantity and quality of suggestions by gender interacted with experience and do not find any compelling evidence for gender differences (see columns 2 and 5 of Appendix Table 1.H.21). While suggestions of lower-experienced females seem to be of significantly lower quality than of lower-experienced males (see column 5), the coefficient is very small in size. This suggests that it's rather experience than gender driving the results.

Effects by disposition of peers and how those differ by workers' own disposition. Having learned that experience is a core determinant of the quantity and quality of suggestions, I now want to explore whether a worker's own experience and the experience of a worker's peers are complements or substitutes. A worker's peers are defined as the group of seating partners per lunch round, who are assigned to be seated opposite, left, right, diagonally left or diagonally right of that worker. Thus, the following analyses only include workers in the lunch group, as they otherwise do not have any assigned seating partners. The experience of peers is defined as the percentage of above-median experienced seating partners, i.e. number of above-median experienced seating partners divided by

[28]The number of processes and the duration to perform each process are recorded in so-called skills matrices. I do not take into account the duration, as the factory prefers and gives priority to workers who can perform multiple operations instead of just a few, which helps meeting line targets. Being able to perform few operations very quickly comes in handy during line balancing, but is less of a priority.

[29]Column 1 and 4 of Appendix Table 1.H.21 shows heterogeneous effects based on gender only, which would imply that female workers submit a significantly lower number of suggestions of lower quality.

TABLE 1.9. INTERACTION EFFECTS OF WORKERS' OWN AND THEIR PEERS' EXPERIENCE

	Quantity		**Quality**	
	(1)	(2)	(3)	(4)
High-exp peers	0.077**	0.12**	0.27*	0.28
	(0.036)	(0.054)	(0.14)	(0.20)
High-exp worker		0.23***		0.77***
		(0.052)		(0.20)
High-exp worker * High-exp peers		-0.11		-0.10
		(0.070)		(0.26)
Observations	2140	2140	2140	2140
R^2	0.05	0.07	0.06	0.07
Control Mean	0.78	0.67	2.97	2.57
Round FE	YES	YES	YES	YES
Floor FE	YES	YES	YES	YES
High-exp peers=High-exp worker		0.05		0.02

Notes: Table 1.9 shows the effect of workers' own and their peers' experience on the quantity and quality of suggestions. Experience is measured by an Anderson (2008) index of the number of factories worked at, grade, tenure in the factory, tenure in the industry and the educational level. "High-exp worker" means that the respective worker has an above-median experience index. "High-exp peers" means that the respective worker is seated around an above-median percentage of high-experienced co-workers, in particular seating partners to the left, right, opposite, diagonally left and diagonally right. At the bottom, I show p-values from comparing the effect of being a high-experienced worker and having high-experienced peers. Regressions exclude observations from workers not assigned to the lunch treatment. They include round and floor fixed effects. Standard errors are clustered at the pair level and shown in parentheses. * p<0.1, ** p<0.05, *** p<0.01

the total number of seating partners.[30] I generate a median of the percentages and use an indicator variable equal to one if the percentage is above the median.

In Table 1.9, I see that the experience of peers matters, but there are no substitution or complementary effects between workers' own and their peers' experience. Both experiences seem to matter individually, although a worker's own experience seem to be more of an important driver (see larger coefficient on own experience compared to the one on peers' experience).

This result raises the question whether there are specific circumstances when the workers' own and their peers' experience are actually complements. Partic-

[30]It is important to define experience of seating partners in percentage terms, as some workers have fewer seating partners than others, e.g. because some seats are at edge of a table or deliberately left empty to mark the seating plan of different units.

ularly in low-trust environments such as in Bangladesh, trustworthiness among discussion partners is considered to be a crucial determinant of exchanging and complementing information. Workers can feel more comfortable and confident to engage into discussions with trustworthy partners, such as friends.[31] The effect of friendship is identified because half of the workers in the lunch group were randomly assigned to be seated opposite a friend. A friend is defined as the co-worker a worker talks to most often about work-related issues, and can be considered as a trustworthy partner. Unlike in the paragraph above, I now define experience of peers as an indicator equal to one if the person seated opposite a worker has an above-median experience index.

I find evidence of a peer-mentor effect: less-experienced workers benefit from being seated opposite a more-experienced friend during lunches. For example, being a low-experienced worker, the quality of suggestions increases by 15% (0.39 from a control mean of 2.55) from being seated opposite a high-experienced partner and by a further 29 percentage points if that partner is also a friend. Being a high-experienced worker, it does not matter whether you are seated opposite a low- or high-experienced partner, regardless of whether that partner is also a friend (see Table 1.H.22). This suggests that specific seating arrangements are only effective for low-experienced workers, once they are seated opposite a high-experienced worker or, even better, a high-experienced friend. Low-experienced workers might have more trustworthy interactions with high-experienced friends, and might feel more aspired and confident to submit their ideas, being 'lectured' by their friends. These patterns are illustrated in Figure 1.7 and corresponding regression results can be found in the Appendix Table 1.H.22.

I further examine whether being seated opposite a new random partner, after two lunch sessions stimulates the generation of ideas, and whether these effects are different depending on whether a worker was seated opposite a friend or non-friend before. The built-in design, in which the seating partners of a third of the workers in the friend and non-friend group are re-randomised after two lunches, allows me to do so. I do not find any significant effects. If at all, being seated opposite a new partner is worse for a worker who has previously been seated opposite

[31] Hahn et al. (2019) find that the individual performance of lower-ability female children in Bangladesh improves once they study with peer friends.



Figure 1.7: Illustration of heterogeneous treatment effects by experience and friendship

Notes: Figure 1.7 illustrates the treatment effects of the 4th column of Table 1.H.22. The labels at the bottom of the figure describe the own experience of the worker first, and the experience of the seating partner and friendship second. For clarity, I also display additional headlines and markings inside the figure. The headlines indicate the workers' own experience. The first three bars show effects for low-experienced workers, whereas the last four bars show effects for high-experienced workers. The markings indicate the seating partners experience, and whether the seating partner is a friend. Bars in grey and pink are low-experienced seating partners, and bars in green and blue high-experienced partners. Circles indicate that the partner is a friend, whereas diamonds indicate that the partner is a non-friend. The horizontal line at 2.55 shows the mean of the control group, which are low-experienced workers, sitting opposite a low-experienced non-friend. 95% confidence intervals are displayed.

a friend, as shown by the negative but insignificant coefficients in columns 2 and 4 of Appendix Table 1.H.23.

Finally, I examine heterogeneous effects by the gender composition during lunches to reinforce the finding that experience is driving results. I calculate the percentage of male peers seated around a worker by dividing the number of men as seating partners by the total number of seating partners. I then construct an indicator variable equal to one for percentages above the median. I do not find any effects of being seated around an above-median percentage of male peers, regardless of the worker's own gender (see column 2 and 5 of Table 1.H.24).[32]

Effects by individual disposition of supervisors

I unpack heterogeneous effects of the huddle treatment by supervisors' individual characteristics. There might be differences in the way supervisors conduct huddles or the way workers perceive huddles. For example, instructions from more senior supervisors might be clearer and come across with higher legitimacy, and hence, serve more as an accelerator for more efficient peer-to-peer interactions.

I restrict regressions to workers assigned to the lunch group only[33], and look at heterogeneous effects by supervisors' average experience, gender and baseline perceived efficacy per line. The experience index of supervisors is similarly constructed as for workers, but does not include grades (as grades are the same for all supervisors). The gender variable measures the percentage of female supervisors per line, and the perceived efficacy index is an Anderson (2008) index of baseline survey questions from supervisors about their perceived likelihood of awards on their line, and perceived motivation, concentration and ability of workers to develop suggestions. On treatment lines with more than one supervisor at baseline, I take line averages.

I do not find any heterogeneous effects for any of those variables on the quantity of suggestions (see columns 1 to 3 of Table 1.10). With respect to the quality of suggestions, I do see heterogeneous effects by the baseline perceived efficacy in

[32]The significant negative effect on the quality of suggestions for female workers in column 5 of Table 1.H.24) disappears once a worker's experience is included in the regression in column 6.

[33]In the absence of the lunch intervention, the huddle treatment had no effects on its own (see Section 1.4.1).

TABLE 1.10. HETEROGENEOUS TREATMENT EFFECTS BY SUPERVISORS' DISPOSITION

	Quantity			Quality		
	(1)	(2)	(3)	(4)	(5)	(6)
Huddle	-0.016	-0.039	-0.044	-0.080	-0.18	-0.36*
	(0.029)	(0.039)	(0.035)	(0.18)	(0.24)	(0.20)
Huddle * Above-median experienced supervisor	-0.024			-0.12		
	(0.052)			(0.32)		
Huddle * Above-median percentage of female supervisors		0.040			0.19	
		(0.048)			(0.29)	
Huddle * Above-median perceived efficacy of supervisors			0.033			0.48
			(0.056)			(0.32)
Observations	2784	2784	2784	2784	2784	2784
R^2	0.13	0.13	0.13	0.08	0.08	0.08
Control Mean	0.25	0.26	0.28	1.21	1.32	1.42
p-value (Huddle=Huddle*Interaction)	0.92	0.35	0.36	0.93	0.48	0.08

Notes: Table 1.10 shows heterogeneous treatment effects of the second managerial intervention by supervisors experience, gender and baseline perceived efficacy. Regressions only include workers assigned to the first managerial intervention, i.e. the lunch treatment. The experience index is an Anderson (2008) of the number of factories worked at, tenure in the factory, tenure in the industry and the educational level. The perceived efficacy index measures supervisors' perceptions about the likelihood of a prize on their line and the likelihood of the implementation of winning suggestions by managers. It also includes an assessment of how motivated and concentrated workers are on their line, as well as how able they are to develop suggestions. I use averages of those variables per line. The percentage of female supervisors is the number of female supervisors per line divided by the total number of supervisors per line. Regressions include floor fixed effects. Standard errors are clustered at the pair level and are shown in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

column 6. Like a self-fulfilling prophecy, it seems as if line huddles from supervisors, who stronger believe in their workers to win and who assess their workers in a more positive light (in terms of motivation, concentration and ability to develop suggestions), are more effective. It suggests that supervisors have a fairly good sense of how their workers will perform rather than that their experience or gender is making a difference.

Learning over time

I finally analyse whether the quantity, quality or variety of suggestions changes over the course of the experiment. Appendix Table 1.H.17 shows that the quantity of suggestions from the lunch group significantly drops between round one and two, but then remains unchanged in subsequent rounds. This is the case for workers assigned to receive both three and four lunches (columns 1 and 2, respectively).

A similar pattern can be observed for the quality of suggestions (see columns 3 and 4) with the only difference that it slightly increases in the final rounds. However, it does not reach the same level as in the first round. Variety also significantly decreases in the third round[34], but only slightly by 7% to 1.33 new suggestions on average per worker. Overall, I can conclude that I cannot find evidence of learning effects as neither the quantity, quality nor variety improves over time.

1.5.3 Does real satisfaction or unspoken dissatisfaction with supervisors explain fewer suggestions on supervisors' skills?

There is a positive and a negative channel through which the second managerial intervention could impact the content of suggestions and decrease the number of suggestions on supervisors' behaviours and skills. On the positive side, the training program may have improved supervisors' skills and/or huddles may have improved workers' perception of their supervisors, and hence, lead to fewer suggestions about their supervisors' skills. I measure real satisfaction through a perceived leadership efficacy index, which contains survey questions about supervisors' abilities. On the negative side, huddles may have deteriorated the voice climate and steered workers away from topics they would otherwise raise. Supervisors might have even manipulated workers to submit suggestions which do not involve negative messages about them in order to protect themselves in front of managers, in line with Friebel and Raith's (2004) theory on skip-level communication.[35] Unspoken dissatisfaction is measured by a perceived voice climate index, which captures the extent to which workers feel comfortable to share their opinions, defined in the literature as 'psychological safety' (Edmondson, 1999; Edmondson and Lei, 2014).

As shown in panel A of Table 1.11, I find no evidence to support the first positive channel of 'real satisfaction'. Workers do not perceive any changes in abilities of their own supervisor or of a typical supervisor in the factory. Further evidence to reinforce this finding is the unchanged mental health index as shown in column

[34]Variety can only be measured if workers submit at least one suggestions, i.e. from round two on-wards and only for the lunch group.

[35]Friebel and Raith (2004) suggest in their theoretical model that skip-level communication, i.e. communication between subordinates and top managers, may induce middle managers to deliberately recruit unproductive subordinates, or abuse their personnel authority in other ways, to protect themselves. Middle managers can have strategic concerns when communication between their subordinates and superiors is too open (Baird, 1977; Lillico, 1972).

TABLE 1.11. REAL SATISFACTION OR UNSPOKEN DISSATISFACTION WITH SUPERVISORS

	Control Mean	Huddle	N
Panel A: Real satisfaction			
Leadership efficacy index	0.03	-0.01	200
		(0.12)	
Ability of own supervisor to encourage operators	-0.03	-0.03	200
		(0.13)	
Ability of typical supervisor to encourage operators	-0.01	-0.03	200
		(0.13)	
Difference in encouragements between own and typical supervisor	-0.01	0.01	200
		(0.13)	
Overall ability of own supervisor	0.12	-0.06	200
		(0.11)	
Overall ability of typical supervisor	0.08	-0.12	200
		(0.12)	
Difference in overall ability between own and typical supervisor	0.03	0.07	200
		(0.10)	
Panel B: Unspoken dissatisfaction			
Voice climate index	0.08	-0.23*	200
		(0.13)	
Sharing of personal issues with superior	0.04	-0.16	200
		(0.14)	
Operators sharing issues with superior	-0.01	-0.06	200
		(0.12)	
Importance of sharing personal issues	0.07	0.03	200
		(0.10)	
Ability to be yourself at work	-0.01	-0.04	200
		(0.11)	
Ability to openly state opinions	0.06	0.01	200
		(0.12)	
Fear of doing something wrong	0.09	-0.30*	200
		(0.17)	
Fear of admitting an error	-0.01	0.03	200
		(0.14)	
Misreporting to supervisor	-0.01	-0.13	200
		(0.15)	
Degree of trust between supervisor and worker (telling a secret)	-0.04	-0.17	200
		(0.15)	
Feeling valued if suggestion made	0.03	-0.11	200
		(0.17)	

Notes: Panel A shows ITT effects of the second managerial intervention on real satisfaction of workers with their supervisors, measured by the leadership efficacy index. Panel B shows ITT effects on unspoken dissatisfaction of workers with their supervisors, measured by a voice climate index. Indices are constructed based on Anderson (2008) and include indented standardised variables beneath, which are coded such that higher values correspond to treatment effects in the same direction. Each row is a separate regression which is based on survey responses from 200 workers. Column 1 shows the mean of workers not assigned to a huddle. Column 2 shows the magnitude of the coefficient with the standard deviation below in parentheses. Column 3 indicates the number of observations N. Floor fixed effects are included in the regressions and standard errors are clustered at the line level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

9 of Table 1.6. In particular, misbehaviour of supervisors - which the majority of suggestions on supervisor skills are about - does not seem to have changed, as indicated by the insignificant estimate on workplace harassment[36] in column 14 of Appendix Table 1.G.13. Furthermore, the self-assessment of supervisors, who have attended the training program, has not changed (see third row and tenth column of Appendix Table 1.H.18).

On the contrary, I can find evidence to support the second negative channel of 'unspoken dissatisfaction', as shown in panel B of Table 1.11. The perceived voice climate of workers significantly deteriorated by 0.23, i.e. around two standard deviations. Workers exposed to line huddles from trained supervisors seem to fear more the consequences of speaking up and feel less safe to state what they think. Interestingly, 30% of all supervisors openly admit in the endline survey to have concerns about workers using the suggestion box to complain about them.[37] Qualitative evidence from discussions with managers also supports these findings. Managers were approached by worrying supervisors who feared negative assessments through workers' suggestions.

Overall, the above described evidence makes me confident to summarize that the huddle treatment comes at the cost of steering workers away from topics they might otherwise raise such as their supervisor's skills. Middle managers, i.e. supervisors – who simultaneously act as principals (to workers) and agents (to higher-level managers) – seem to be reluctant to foster free upward information flows from workers to higher-level managers as these flows might involve negative messages about them.

1.5.4 Why are workers more satisfied?

As outlined in Section 1.4.2, I find that the lunch treatment leads to higher job satisfaction. It remains to be explored whether this results is driven by perceived actual changes in the working conditions or by the act of exerting one's voice regardless of actual changes in firm policies or managers' behaviour. I consider two aspects

[36] Workplace harassment is measured by the frequency of uncomfortable situations at work and of abusive language and shouting from supervisors.

[37] This percentage is twice as high for untrained supervisors than for trained supervisors, but the difference is not significant.

of working conditions which I expect to change the most as a result of structured peer-to-peer lunches: Communication practices and workplace culture.

In Table 1.12 I show that perceived working conditions do not change by the lunch intervention. The general communication index, which measures the frequency and topics of discussions between workers and supervisors, and the raising suggestions index, which measures the frequency of verbally raising suggestions to supervisors or managers on the production floor, remain unchanged. Only communication about the suggestion box itself is significantly increasing by 0.18, which was to be expected. The voice climate index, which measures the extent to which workers feel free to share their opinions, the caring culture index, which measures feelings at the workplace, and the mindset about authority index, which measures the extent to which workers perceive conflict between workers and managers, do not change.

I argue that the act of exerting one's voice makes workers more satisfied, which is in line with what has been causally explained in the literature (Adhvaryu et al., 2019).[38] Being aware of endogeneity, significantly positive heterogeneous effects on job satisfaction by an above-median number of submissions are at least indicative to support this argument. Column 1 in Appendix Table 1.H.25 shows that workers in the lunch group who have actually raised their voice seem to be more satisfied than those who did not, which is not due to higher experience (see column 2). This suggests that job satisfaction is driven by the actual act of exerting one's voice, rather than by the mere opportunity to raise one's voice or perceived changes in working conditions.[39]

[38] Adhvaryu et al. (2019) show that giving randomly selected workers the opportunity to participate in a feedback survey after a disappointing wage hike reduced turnover. As participants were told that the firm would not receive the feedback provided, they can conclude that the act of being asked for one's opinion is driving the result rather than possible changes in firm policy or managers' behaviour.

[39] In my theoretical framework described in Appendix Section 1.A, this result corresponds to changes in the intrinsic motivation of workers to generate ideas, $M(e_i; \lambda_i, N_{ij})$, which captures the fact that workers derive utility from thinking about ideas regardless of winning, for example through enjoyment from being listened to and cared for or inner satisfaction from thinking about ideas individually or jointly.

TABLE 1.12. EFFECTS ON COMMUNICATION PRACTICES AND WORKPLACE CULTURE

	Control Mean	Lunch	N
Overall index: Communication practices	-0.02	-0.02	200
		(0.05)	
General communication index	-0.01	0.06	200
		(0.15)	
Raising suggestions index	-0.10	-0.00	200
		(0.07)	
Suggestion box communication index	-0.06	0.14	200
		(0.13)	
Suggestion box communication index (all workers)	0.05	0.18***	1164
		(0.06)	
Overall index: Workplace culture	0.04	-0.03	200
		(0.14)	
Voice climate index	0.01	0.05	200
		(0.13)	
Caring culture index	0.04	-0.07	200
		(0.11)	
Mindset about authority index	0.04	-0.13	200
		(0.14)	

Notes: Table 1.12 shows ITT effects of the first managerial intervention on workers' perceived communication practices and workplace culture. Overall indices include indented indices beneath, which are constructed based on Anderson (2008). Individual index components can be seen in Appendix Figure 1.E.9. Each row is a separate regression. Column 1 shows the mean of workers not assigned to a lunch. Column 2 shows the magnitude of the coefficient with the standard deviation below in parentheses. Column 3 indicates the number of observations N. It is higher for the "Suggestion box communication index (all workers)" as corresponding survey questions were asked to all workers at endline. Floor fixed effects are included in the regressions and standard errors are clustered at the pair level. * p<0.1, ** p<0.05, *** p<0.01

1.6 Conclusion

In this paper, I examine whether changing management practices to encourage bottom-up innovation leads to beneficial information flows from production workers on lower hierarchical levels to higher-level management with the power to implement ideas. To do this, I set up a suggestion system in a Bangladeshi garment factory, in which the best three ideas per floor are financially rewarded and the winners' ideas displayed for recognition. I conduct a randomised controlled trial with two cross-randomised managerial interventions targeted at reducing costs of effort and time, social costs, and the combination of both costs. First, targeted at costs of effort and time, I followed Sandvik et al. (2020) and separated out time for randomly selected workers to engage in discussions about ideas under the help of external facilitators during free peer-to-peer lunch sessions ("lunch treatment"). Second, targeted at social costs, I randomly assigned supervisors to receive a two-hour training program on worker communication methods and to give verbal encouragements to workers in the form of monthly three-minute line huddles ("huddle treatment"). It is the first study to provide empirical evidence on the effects of voice-enhancing managerial practices to address impediments to upward flows of ideas from production workers to top managers in low-income countries.

My paper shows that the bottom-up channel of ideas can be fostered through structured peer-to-peer lunches, and simultaneously lead to improvements in worker well-being. I find that workers assigned to the lunch treatment submit significantly more suggestions than the non-lunch group (1.809 vs 48 suggestions over the intervention period of three months), but the quality is 20% lower than of those from the non-lunch group. This suggests positive self-selection of high quality submissions from the non-lunch group. On the contrary, the lunch group submits suggestions of much larger breadth, covering diverse topics. Besides, I find that the lunch treatment significantly improves workers' job satisfaction, but I do not find any short-term effects on firm performance. However, I see lower absenteeism on assigned lunch treatment days, which shows that workers enjoy being involved in the suggestion process and receiving a free lunch gift from the firm. Involving middle managers does neither have any effects on the quantity, quality and variety of suggestions, nor on firm performance and worker well-being, and

even comes at the cost of reducing the number of suggestions on supervisors' skills and behaviours.

Overall, the huddle treatment is not beneficial. In contrast, the lunch treatment has clear benefits. I provide three explanations in favour of trading off quality against quantity, i.e. to get more suggestions of varying quality from the lunch group rather than fewer suggestions of high quality from the non-lunch group. First, the lunch treatment makes workers better off regardless of the quality of submitted suggestions, which could eventually lead to reduced turnover in the long-run. Second, the lunch treatment is a cost-effective intervention in improving worker well-being; it is not costly to evaluate suggestions, and 88% of the suggestions judged to be best (blind to treatment) were from the lunch group. Third, the lunch treatment enables the firm to get a feeling for what workers think and care about through the larger breadth of submitted topics. Implementing suggestions on those topics could eventually lead to productivity improvements in the long-run. In fact, some workers already happily noticed the implementation of a few suggestions at endline, such as the instalment of new fans or the repair of broken floor tiles.

I claim that the lunch intervention is successful because it makes it easier for workers to submit ideas and increases their aspirations. Workers' perceived abilities do not seem to have changed, but they are indeed an important determinant of the quality and quantity of suggestions at the intensive margin. Further, I find evidence of a peer-mentor effect: less-experienced workers benefit from being seated opposite a more-experienced friend during lunches. Besides, I find that the lower number of suggestions on supervisor skills from workers assigned to the huddle treatment can be explained by a perceived deterioration in the voice climate, in line with Friebel and Raith's theory on skip-level communication (Friebel and Raith, 2004). Finally, I find that the act of exerting voice drives job satisfaction rather than actual perceived changes in working conditions, confirming the results in the literature (Adhvaryu et al., 2019).

My findings raise several important directions for future research. First, this research highlights impediments to upward flows of ideas. Having found that aspirations and confidence are main drivers for idea generation, more targeted follow-

up interventions could be conducted which directly address these channels. For example, future research could look at the effects of role models in changing workers' aspirations and confidence, e.g. by varying the characteristics of the winners whose ideas and names are announced and displayed.

Second, future research could more fully investigate whether effects on job satisfaction are persistent and eventually translate into reductions of turnover rates. Besides, it would be interesting to analyse how productivity is affected in the long-run through successful implementation of ideas.

Third, I do find that upward communication is meaningful for the firm, but at the end of the experiment, it became evident that workers lost interest into the suggestion box and into engaging in discussions with co-workers over lunch. In the final lunch round, workers tried to just pick up their lunch and go away, signalling that the managerial practice lost its strength in fostering idea generation. It remains interesting to analyse how the idea generation process can be made more sustainable and incorporated over longer periods, which would then also allow to more closely analyse learning over time.

Fourth, a critical question is how the involvement of middle managers can be made more effective, e.g. in the form of non-monetary or monetary incentive structures. The setting of this study did not allow to change monetary incentives for supervisors, but discussions with industry experts suggested that it might have been a lever for a more successful involvement of supervisors.

Overall, this paper has shown that it is indeed possible and effective to improve upward flows of ideas of workers, who have never been asked for their opinion beforehand, through managerial interventions, but there remains a lot more research to be done on further targeted interventions to strengthen a culture of bottom-up innovation.

Acknowledgements I am grateful to my supervisors, Christopher Woodruff and Simon Quinn, for their insightful feedback, advice and support. I also thank Sanjay Jain, Martin Williams, Meg Meyer, Lukas Hensel, Daniela Scur, Thomaz Teodorovicz, Florian Englmaier, Kathryn Shaw and seminar participants at the Cedefop, Eurofound and IZA Conference on Workplace and Management Practices, the Oxford's CSAE Workshop and Oxford's Firms, Labour and Organisations Research Group for their comments and suggestions. I am grateful to the field managers and enumerator team at the Innovations for Poverty (IPA) office in Bangladesh for their hard work on this project. I thank the the participating factory for their support and cooperation. I am thankful for generous financial support of this research from the PEDL Special Exploratory Research Grant for PhD Students (Private Enterprise Development in Low-Income Countries). Ethical approval was obtained on the 26th of November 2019 from the Oxford Department of Economics' Departmental Research Ethics Committee (DREC) with reference ECONCIA19-20-24.

Appendices of Chapter 1

1.A Theoretical framework

In the following, I present a general theoretical framework for understanding employee voice, which has been discussed non-formally to date.

1.A.1 Set-up

Workers on the production floor can exert effort e to submit their ideas about what they would like to improve inside the factory. For simplicity, I assume that workers think about ideas either individually ($N_{ij} = 0$) or with a co-worker j ($N_{ij} = 1$). The co-worker can either be a friend ($G_{ij} = 1$) or a non-friend ($G_{ij} = 0$). This set-up allows us to look at the drivers of idea generation among three different groups of workers:

- Individuals: $N_{ij} = 0$
- Friend pairs: $N_{ij} = 1$ and $G_{ij} = 1$
- Non-friend pairs: $N_{ij} = 1$ and $G_{ij} = 0$

I further assume that all workers have the possibility to win a prize for the best submitted idea. Workers who think about ideas with a co-worker have to share the prize with their sparring partner, if they submit the same idea.[40]

1.A.2 Utility Function

Individual i derives utility U from submitting ideas. On the one hand, worker i receives utility if she wins the tournament (benefit of winning B), which depends on the probability of winning P. On the other hand, worker i also receives utility from intrinsic motivation independently of the probability of winning (effort incentive M). In fact, the policy of asking workers to generate ideas may have motivated them to exert effort independently of the effect of their effort on the probability of winning a prize. Finally, worker i faces costs of effort C to be deducted from the benefits. Hence, the utility function is specified as below:

$$\begin{aligned} U_i(e_i, e_j, e_{-ij}; \theta_{ij}, \delta_i, N_{ij}) = P(e_i, e_j, e_{-ij}; \theta_{ij}, N_{ij}) \times B(v_f, v_i, v_a; N_{ij}) \\ + M(e_i; \lambda_i, N_{ij}) - C(e_i) \end{aligned} \tag{1.5}$$

where

- $P(e_i, e_j, e_{-ij}; \theta_{ij}, \delta_i, N_{ij}) \in (0, 1)$ is the **probability of winning a prize** which I define as a function of efforts, i.e. individual i and paired worker's j own

[40] In reality, the factory did not award shared prizes, but this was the policy they announced in case the winning idea was submitted by someone else as well.

efforts and the expected average level of effort of all workers other than the individual and paired worker e_{-ij} (called $\bar{e}$). Due to the competitive tournament structure (only the best ideas win), I assume that the expected effort of everyone else is a substitute on a global level, i.e. higher $\bar{e}$ lowers the probability of winning for individual/paired workers. Hence, $\bar{e}$ enters the probability function negatively. The intensity of the decline of worker i's effort in response to changes in $\bar{e}$ (degree of strategic substitutability) is captured by the **global interaction parameter** δ_i. On a local level within paired workers, I assume that worker j's effort can be a substitute or a complement, i.e. worker j's effort can lower or increase the probability of winning of worker i. In other words, workers can be less productive (substitutes) or more productive (complements) jointly. The change in effort of worker i in response to changes of e_j (degree of strategic substitutability or complementarity) depends on the **local interaction parameter** θ_{ij}. Both interaction parameters are described in more detail in the following:

  - δ_i is defined as the **global interaction parameter** of worker i. For simplicity, I assume $\delta_i = \delta_j$. It captures the intensity to which worker i decreases her own effort given the expected average of everyone else's effort. It depends on the worker's belief about her own probability of winning. If the worker strongly believes that she wins, she will be less inclined to adjust her effort levels to the expected average (small δ_i). If the worker is sceptical of winning (large δ_i), the worker will be inclined to exert less effort the higher the expected average effort (due to the competitive tournament structure).
  - θ_{ij} is defined as the **local interaction parameter** between worker i and paired worker j. It captures the intensity to which worker i increases or decreases her own effort given the effort of worker j. It depends on how well paired workers interact, communicate and coordinate with each other. Factors such as degree of friendship or differences in educational levels might play a role. For example, if worker j works harder and is a good friend of worker i, the latter mights also want to work harder. A possible reason could be more trust among friends leading to better interaction patterns, e.g. they could be less ashamed to screen out less good ideas quicker, to say different perspectives, or to discover "blind spots" in thinking more easily. As another example, if worker j has a higher educational level than worker i leading to less common understanding, worker i might want to free-ride and not exert any effort.

- $B(v_f, v_i, v_a)$ is the **benefit of winning a prize** where
  - v_f: **Extrinsic benefit** of winning a prize, i.e. financial reward
    - α captures the share to which the financial reward is divided among pairs
      - $\alpha = 1$: Ideas submitted by individuals

· $\alpha = 0.5$: Ideas submitted by pairs. Workers are facing a loss of the extrinsic benefit v_f by 50%, because they have to share the monetary reward with their paired co-worker.

– v_i: **Intrinsic benefit of individual/joint achievement**, e.g. social approval and recognition (by displaying the winner's name), feeling accomplished about oneself/pair, doing something good for the employer and co-workers, etc.

∗ β captures degree of preference for individual/joint consumption. It depends on the friendship between worker and co-worker, but also on differences in education (e.g. a worker might not enjoy joint consumption if the co-worker is not as educated)

· $\beta > 1$: worker prefers winning jointly/sharing the experience; higher utiliy in case of joint consumption

· $\beta = 1$: indifference between joint and individual consumption

· $\beta < 1$: worker prefers winning alone; lower utility in case of joint consumption

– v_a: **Intrinsic benefit of altruistic behaviour**, e.g. caring about the partner to win and to receive a financial benefit

∗ γ captures the degree of altruism for partner

· $\gamma > 1$: care about partner to win

· $\gamma = 1$: indifference

· $\gamma < 1$: dislike partner to win

- $M(e_i; \lambda_i, N_{ij})$ is the **intrinsic motivation of generating ideas** (regardless of winning), e.g. enjoyment from being listened to and cared for, inner satisfaction from thinking about ideas individually or jointly, joy of doing something good for the employer and co-workers, etc. It is a function of worker i's effort. λ_i captures the size of the intrinsic motivation effect and raises the marginal return to effort, and hence the equilibrium effort.

- $C(e_i)$ is the **cost of effort**, which is a function of individual i's effort to capture differences in individual abilities, i.e. the cost to think about ideas, to draft them in a way managers can understand and reward and to hand them in (including opportunity costs of time). I assume regularity conditions to hold $C'() > 0, C''() > 0$, $lim_{e\to\infty} C'(e) = \infty$ guaranteeing the existence of a unique solution.

In the following, I define the probability and cost functions directly in terms of effort assuming concrete functional forms. I normalize all parameters and effort levels to have a valid probabilty density function. The form of the probabilty function was chosen such that utilitiy is always increasing with the worker's own effort, but can be decreasing or increasing with the co-worker's effort depending

on the model parameters. I have selected a simple additive form for the benefit function ($B = \alpha v_f + \beta v_i + \gamma v_a$).
The nested utility function is as follows

$$U = [\frac{1}{3}(1 - \delta_i \bar{e})e_i + \frac{1}{3}N_{ij}((1 - \delta_j \bar{e})e_j + \theta_{ij}e_i e_j)]B + \lambda_i e_i - \frac{e_i^2}{2} \tag{1.6}$$

with e_i, e_j and $\bar{e}$, δ_i, δ_j, λ_i and $B \in [0, 1]$, and $\theta_{ij} \in [-1, 1]$.

1.A.3 Optimization

Individual i maximizes utility function U with respect to its own effort e_i:

max U w.r.t. e_i

$$\frac{\partial U}{\partial e_i} = \frac{1}{3}B[(1 - \delta_i \bar{e}) + N_{ij}\theta_{ij}e_j] + \lambda_i - e_i \tag{1.7}$$

The optimal response functions and how they change depending on parameter values are graphically depicted in the appendix. The second partial derivative with respect to e_j is

$$\frac{\partial U}{\partial e_i \partial e_j} = \frac{1}{3}BN_{ij}\theta_{ij} \tag{1.8}$$

Whether efforts of worker i and co-worker j are strategic complementarities or strategic substitutes depends on the value of parameter θ:

- $\frac{\partial U}{\partial e_i \partial e_j} < 0$ if $\theta_{ij} < 0$: **Strategic substitutes.** Worker i exerts less effort if worker j exerts more effort, i.e. there is a low level of communication and coordination. Note that worker j could also free-ride and not exert any effort if worker i is working ($\theta = 0$)
- $\frac{\partial U}{\partial e_i \partial e_j} > 0$ if $\theta_{ij} > 0$: **Strategic complements**. Worker i exerts more effort if worker j exerts more effort, i.e. there is a high level communication and coordination among workers.

Assuming the same global interaction parameters ($\delta_i = \delta_j$) and the same size of the intrinsic motivation effect ($\lambda_i = \lambda_j$) per pair, I get the following Nash equilibrium level of effort (proof see Appendix Section 1.A.7):

$$e_i^\star = \frac{B(1 - \delta_i \bar{e}) + 3\lambda_i}{3 - N_{ij}B\theta_{ij}} \tag{1.9}$$

Using an additive functional form for B yields

$$e_i^\star = \frac{(\alpha v_f + \beta v_i + \gamma v_a)(1 - \delta_i \bar{e}) + 3\lambda_i}{3 - N_{ij}(\alpha v_f + \beta v_i + \gamma v_a)\theta_{ij}} \tag{1.10}$$

This equilibrium equation indicates that worker i does not exert any effort if $\delta_i \bar{e} = 1$ and $\lambda_i = 0$. The optimal response functions and how the equilibrium efforts change if for example θ or B increase are illustrated in figures 1.A.6 and 1.A.6 in case of strategic complements, and in Figure 1.A.6 in case of strategic substitutes.

For each group the parameters are specified as below.

1. **Individuals** ($N_{ij} = 0$): $e_j = 0$, $\alpha = 1$, $\beta = 1$, $v_a = 0$, $\delta_i = \delta_C$ and $\lambda_i = \lambda_C$
   Individuals are not paired with anyone ($e_j = 0$). They receive the full financial reward v_f ($\alpha = 1$) and they can only receive the intrinsic benefit from individual consumption, not from joint consumption ($\beta = 1$). Also they cannot receive any benefit from altruistic behaviour ($v_a = 0$). The global interaction parameter is δ_C and the intrinsic motivation effect has the size λ_C.

2. **Friend pairs** ($N_{ij} = 1$, $G_{ij} = 1$): $e_j > 0$, $\theta_i = \theta_F$, $\alpha = 0.5$, $\beta = \beta_F$, $\gamma = \gamma_F$, $\delta_i = \delta_F$ and $\lambda_i = \lambda_F$
   Individuals are paired with a friend who exerts effort e_j. The level of communication and coordination among friends is captured by the local interaction term θ_F. Individuals have to share the prize with their partner ($\alpha = 0.5$). They have consumption preferences β_F and their degree of altruism takes the value γ_F. The global interaction parameter is δ_F and the intrinsic motivation effect has the size λ_F.

3. **Non-friend pairs** ($N_{ij} = 1$, $G_{ij} = 0$): $e_j > 0$, $\theta_i = \theta_{NF}$, $\alpha = 0.5$, $\beta = \beta_{NF}$, $\gamma = \gamma_{NF}$, $\delta_i = \delta_{NF}$ and $\lambda_i = \lambda_{NF}$
   Individuals are paired with a non-friend who exerts effort e_j. The level of communication and coordination among non-friends is captured by the local interaction term θ_F. Individuals have to share the prize with their partner ($\alpha = 0.5$). They have consumption preferences β_{NF} and their degree of altruism takes the value γ_{NF}. The global interaction parameter is δ_{NF} and the intrinsic motivation effect has the size λ_{NF}.

I assume that the parameters fulfill the following criteria:

- $\theta_F > \theta_{NF}$: The local interaction parameter is lower for friends than for non-friends, with a lower parameter indicating a higher level of communication and coordination, e.g. due to a higher level of trust.

- $\delta_F < \delta_{NF}$: The global interaction parameter is lower for friend than the non-friend pairs, indicating that the their level of effort is less susceptible to changes in the average level of effort, e.g. because they work well together and think that their probability of winning is high.

- $\beta_F > \beta_{NF}$: The preference for joint consumption for friends is higher than for non-friends, e.g. because it is more rewarding to share the experience of winning with someone a worker cares about.
- $\gamma_F > \gamma_{NF}$: The degree of altruism for a friend is higher than for a non-friend, e.g. because a worker would care more about someone she knows well to win the prize.
- $\lambda_F > \lambda_{NF}$: The size of the intrinsic motivation effect for a friend is higher than for a non-friend, e.g. because thinking about ideas with a friend is much more fun than with a non-friend (or alone).

I can now plug in the parameters to find the optimal level of effort per group.

Optimal effort of individuals

$$e_i^\star = \frac{1}{3}B_C(1-\delta_C\bar{e}) + \lambda_C = \frac{1}{3}(\alpha v_f + \beta v_i + \gamma v_a)(1-\delta_C\bar{e}) + \lambda_C \tag{1.11}$$

Optimal effort of friend pairs

$$e_i^\star = \frac{B_F(1-\delta_F\bar{e}) + 3\lambda_F}{3 - B_F\theta_F} = \frac{(\alpha_F v_f + \beta_F v_i + \gamma_F v_a)(1-\delta_F\bar{e}) + 3\lambda_F}{3 - (\alpha_F v_f + \beta_F v_i + \gamma_F v_a)\theta_F} \tag{1.12}$$

Optimal effort of non-friend pairs

$$e_i^\star = \frac{B_{NF}(1-\delta_{NF}\bar{e}) + 3\lambda_{NF}}{3 - B_{NF}\theta_{NF}} = \frac{(\alpha_{NF} v_f + \beta_{NF} v_i + \gamma_{NF} v_a)(1-\delta_{NF}\bar{e}) + 3\lambda_{NF}}{3 - (\alpha_{NF} v_f + \beta_{NF} v_i + \gamma_{NF} v_a)\theta_{NF}} \tag{1.13}$$

1.A.4 Predictions

In this section, I want to test the following two predictions:

1. **Friendship effect** (effect of being paired with a friend): The quality and quantity of suggestions is higher for friend pairs than for non-friend pairs.
2. **Social effect** (effect of being paired with a co-worker): The quality and quantity of suggestions is higher for pairs than for individuals.

These two predictions can be mathematically expressed as:

1. $e_i^\star\big|_{N_{ij}=1,G_{ij}=1} > e_i^\star\big|_{N_{ij}=1,G_{ij}=0}$
2. $e_i^\star\big|_{N_{ij}=1} > e_i^\star\big|_{N_{ij}=0}$

Friendship effect

Inserting optimal efforts into prediction 1) yields

$$\frac{B_F(1-\delta_F)+3\lambda_F}{B_{NF}(1-\delta_{NF})+3\lambda_{NF}} > \frac{3-B_F\theta_F}{3-B_{NF}\theta_{NF}} \tag{1.14}$$

Given that B_F is greater than B_{NF} (as $\beta_{NF} < \beta_F$ and $\gamma_{NF} < \gamma_F$), δ_F is smaller than δ_{NF} and λ_F is greater than λ_{NF}, the LHS is always larger than one. This would also the case even if δ_F were equal to δ_{NF}. Given that $\theta_F > \theta_{NF}$, the RHS is always smaller than one. Hence, prediction 1) holds independent of whether efforts of non-friend pairs are complements (weaker than of friend pairs) or substitutes.

Social effect

As I just showed that the optimal effort of friend pairs always higher than for non-friend pairs, I just have to show that the effort for non-friend pairs is higher than for individuals to proof prediction 2: $e_i^\star\big|_{N_{ij}=1,G_{ij}=1} > e_i^\star\big|_{N_{ij}=1,G_{ij}=0} > e_i^\star\big|_{N_{ij}=0}$

Inserting optimal efforts of the non-friend pairs and of individuals into prediction 2 yields

$$\frac{3}{3-B_{NF}\theta_{NF}} > \frac{B_C(1-\delta_C)+3\lambda_C}{B_{NF}(1-\delta_{NF})+3\lambda_{NF}} \tag{1.15}$$

If efforts of non-friend pairs are complements $\theta_{NF} > 0$, the LHS is always greater than one. If efforts of non-friend pairs are substitutes $\theta_{NF} < 0$, the LHS is always smaller than one.

The value of the RHS depends on the size of the benefits (i.e. benefit of winning B and intrinsic motivation effect λ) of individuals relative to non-friend pairs, and the size of the global interaction parameter δ (i.e. how much effort changes in response to changes in the average level of effort): $(v_f+v_i)+\frac{3\lambda_C}{1-\delta_C} \gtrless (\frac{v_f}{2}+\beta_{NF}v_i+\gamma_{NF}v_a)+\frac{3\lambda_{NF}}{1-\delta_{NF}}$ which can be rewritten as $\frac{v_f}{2} \gtrless (\beta_{NF}-1)v_i+\gamma_{NF}v_a+\frac{3\lambda_{NF}}{1-\delta_{NF}}-\frac{3\lambda_C}{1-\delta_C}$. It shows that the result depends on the size of the loss in financial benefit from sharing the prize relative to the gain in benefits from working with a co-worker (joint consumption/winning, altruistic behaviour and intrinsic motivation) as well as the effort adjustments to the average level of effort. If efforts of non-friend pairs are substitutes (LHS < 1) indicating that the interaction/coordination between worker and co-worker does not work well, the gain in benefit from working with the co-worker must be sufficiently large (e.g. $\beta_{NF} > 1$, $\gamma_{NF} > 1$, and $\lambda_{NF} > \lambda_C$) and/or the global interaction parameter sufficiently small (e.g. $\delta_{NF} < \delta_C$) for the non-friend pairs to exert more effort (so that 1 > LHS = $e^*_{NF} \gg$ RHS = e^*_C). If efforts are complements and the probability of winning increases in the co-workers effort j (LHS > 1), the gains in benefit from working with a co-worker do not need to be as large for the non-friend pairs to exert more effort (so that LHS = e^*_{NF} > 1 > RHS = e^*_C).

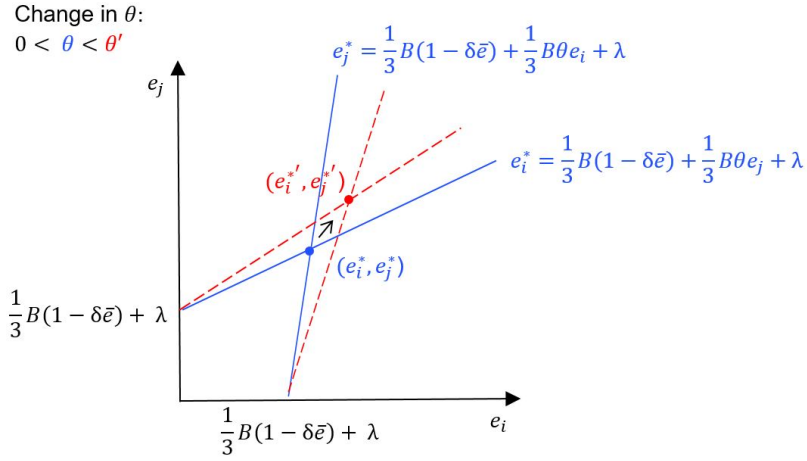
1.A.5 Validity of model assumptions

It is important to acknowledge that the average effort of everybody else $\bar{e}$ is really endogenous as the experiment itself changes the average level of idea generation on the production floor. I argue that in this model the endogeneity of $\bar{e}$ does not matter, but rather the assumption that $\bar{e}$ is the same for all groups of workers. This model allows for intragroup comparisons of three groups of workers (friend pairs, non-friend pairs or individuals) on one production floor under a tournament structure. However, I must be careful when extrapolating from this setting to other settings with a different $\bar{e}$.

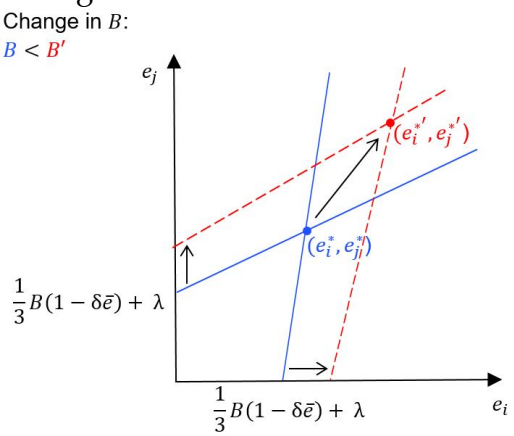
1.A.6 Graphical display of strategic complements and substitutes

Strategic complements ($\theta > 0$)

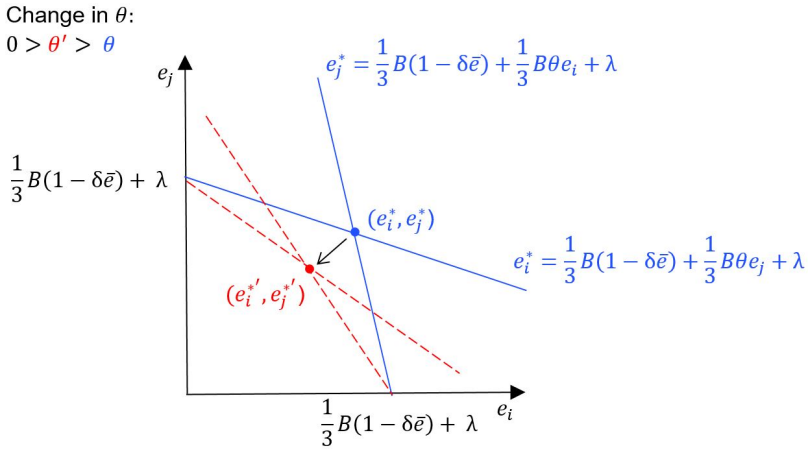
Change in θ



Change in B



Strategic substitutes ($\theta < 0$)



1.A.7 Mathematical Appendix: Proof of equation 1.9

$$\frac{\partial U}{\partial e_i} = \frac{1}{3}B[(1-\delta_i\bar{e}) + N_{ij}\theta_{ij}e_j] + \lambda_i - e_i \tag{1.16}$$

$$\frac{\partial U}{\partial e_j} = \frac{1}{3}B[(1-\delta_j\bar{e}) + N_{ij}\theta_{ij}e_i] + \lambda_j - e_j \tag{1.17}$$

Plugging (1.17) into (1.16) with $\delta_i = \delta_j$ and $\lambda_i = \lambda_j$ yields

$$\frac{1}{3}B[(1-\delta_i\bar{e}) + N_{ij}\theta_{ij}(\frac{1}{3}B[(1-\delta_i\bar{e}) + N_{ij}\theta_{ij}e_i] + \lambda_i)] + \lambda_i = e_i$$

$$\frac{1}{3}B(1-\delta_i\bar{e})(1+\frac{1}{3}N_{ij}B\theta_{ij}) + \lambda_i(1+\frac{1}{3}N_{ij}B\theta_{ij}) = e_i(1-N_{ij}^2\frac{1}{9}B^2\theta_{ij}^2)$$

$$e_i = \frac{(\frac{1}{3}B(1-\delta_i\bar{e}) + \lambda_i)(1+\frac{1}{3}N_{ij}B\theta_{ij})}{(1-N_{ij}^2\frac{1}{9}B^2\theta_{ij}^2)}$$

$$e_i = \frac{(\frac{1}{3}B(1-\delta_i\bar{e}) + \lambda_i)(1+\frac{1}{3}N_{ij}B\theta_{ij})}{(1+\frac{1}{3}N_{ij}B\theta_{ij})(1-\frac{1}{3}N_{ij}B\theta_{ij})}$$

$$e_i^\star = \frac{B(1-\delta_i\bar{e}) + 3\lambda_i}{3 - N_{ij}B\theta_{ij}}$$

1.B Determination of line location of workers

The production lines of workers were determined based on daily line movement data from the IE department. However, only 40% of all days per worker are covered during the intervention period. 159 out of 1599 workers are not included at all and some days are missing for individual workers. Data of one unit is completely missing for January and March 2020. 40% of the 159 workers without any information work in that particular unit. Besides, more than one line is recorded for some workers. Due to this incompleteness, the dataset has to be complemented with survey and intervention data. I complement the data with the following approach. First, I use the line code as indicated on suggestion papers for the days at which workers submitted their papers. Second, I 'fill up' line codes with the last available observation assuming that workers continued on the line mentioned last. If the last entry contained several lines, I assume that workers continued on the line which appears next. Mostly, workers were working on two lines just for one day, and continued on the second line going forward. Third, I use line codes from the first available date in February 2020 for all workers from one unit for which the factory did not record any data in January 2020. Fourth, I use line codes from the endline data for missing lines in March 2020. Fifth, I use line codes from the baseline data for the 159 workers without any observations in the line movement data. The distribution of the different data sources to determine each worker's line per day is shown in Figure 1.B.1.

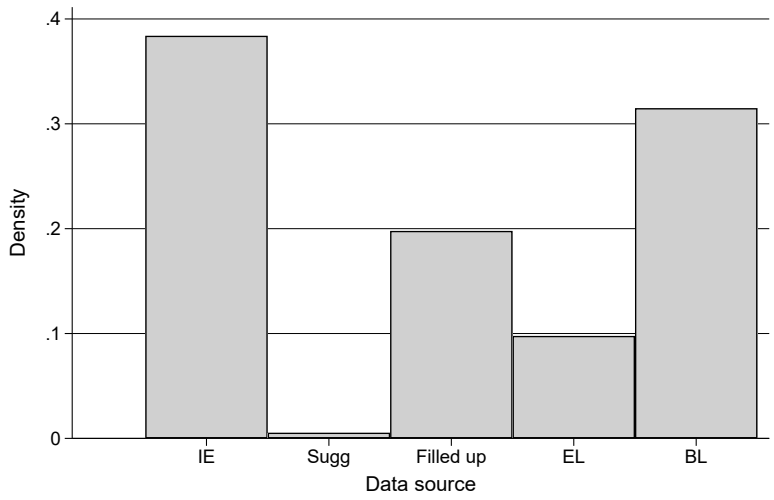


Figure 1.B.1: Data sources for line location of workers

Notes: Graph 1.B.1 shows the distribution of the different data sources to determine the line location of workers. "IE" means line movement data from Industrial Engineers. "Sugg" means line data from the suggestion papers (only if no IE data available). "Filled up" is missing line movement data which is complemented by available data from dates before or after. "EL" indicates that the line code is taken as provided in the endline survey, and "BL" as provided in the baseline survey (if no IE or Suggestion data available).

1.C Implementation details and documentation

TABLE 1.C.1. INTERVENTION DATES

Panel (a): Dates of peer-to-peer lunch sessions for treatment group (lunch intervention)	
Round 1: First lunch - Subgroup 1	26.12.2019
Round 1: First lunch - Subgroup 2	01.01.2020
Round 1: First lunch - Subgroup 3	08.01.2020
Round 1: First lunch - Subgroup 4	15.01.2020
Round 2: Second lunch - Subgroup 1	23.01.2020
Round 2: Second lunch - Subgroup 2	30.01.2020
Round 2: Second lunch - Subgroup 3	05.02.2020
Round 2: Second lunch - Subgroup 4	12.02.2020
Round 3: Third lunch - Subgroup 1	24.02.2020
Round 3: Third lunch - Subgroup 2	27.02.2020
Round 3: Third lunch - Subgroup 3	03.03.2020
Round 3: Third lunch - Subgroup 4	11.03.2020
Round 4: Fourth lunch - Subgroup 1	15.03.2020
Panel (b): Dates of take-away lunch distribution for control group (lunch intervention)	
Take-away lunch (in round 2) - Subgroup 1	21.01.2020
Take-away lunch (in round 2) - Subgroup 2	26.01.2020
Take-away lunch (in round 2) - Absent workers 1	03.02.2020
Take-away lunch (in round 2) - Absent workers 2	12.02.2020
Panel (c): Dates of training and sms reminders (huddle intervention)	
Motivational training	06.01.2020
Reminder I (2 sms)	13.01.2020
Reminder II (2 sms)	08.02.2020
Technical training	04.03.2020
Reminder III (2 sms)	09.03.2020

First sms: "Dear Supervisors, Thank you for participating in the suggestion box project. Please remember to gather all workers on your line together for 5 minutes and encourage them to submit ideas. Please make sure to meet between XX.XX.2020 and XX.XX.2020 [*two days after reminder I and three days after reminder II and III*]. Also remember to fill out the meeting form that you can get from the floor IE."
Second sms: "Dear Supervisor, You are one of the only supervisors who have not met your workers yet for the suggestion box project. Please be reminded to gather all workers on your line together for 5 minutes and encourage them to submit ideas. Please make sure to meet by tomorrow and remember to fill out the meeting form that you can get from the floor IE. Many thanks."

Figure 1.C.2: SMS texts for supervisors

IDEA BOX (1 idea per paper)

Please write legibly.

Name		**Worker ID**	**Line code**
Department	**Unit**	**Designation**	**Date** ____/____/20____
Problem	**Idea to solve problem**		

Area of improvement – Please tick the box(es) that match your idea

☐ Absenteeism ☐ Defects ☐ Efficiency ☐ Machines ☐ Motions/Waiting time ☐ Production flows/layout

☐ Safety ☐ Shipment ☐ Skills ☐ Wastage ☐ Work environment ☐ Other (please specify): _____

Figure 1.C.3: Suggestion paper

Suggestion Box Line Huddle

Date ……. /……. /2020

Duration of huddle in minutes ……

Your name ……

Your ID ……

Your line and unit ……

No. of line workers ……..

No. of present workers at huddle……..

Approx. no. of ideas from last week ……

Topics discussed

☐ Remind workers of idea box ☐ Motivate workers ☐ Provide help and answer questions ☐ Explain how to fill out idea paper

Which questions were you asked?

Signature Supervisor …… Signature In-charge ……

Figure 1.C.4: Template for line supervisors to track conduct of huddles

Suggestion Box – Evaluation Committee
Nomination form

Please select your unit
☐ Unit 1 ☐ Unit 2 ☐ Unit 3
☐ Unit 4 ☐ Unit 5 ☐ Unit 6

Date ……. /……. /2020

What are the best three suggestions?

NO.	IDEA	REASONS FOR SELECTION
#1. Gold		
#2. Silver		
#3. Bronze		

Name, Signature

IE ……, …… Sewing ……, ……

Quality ……, …… Sewing ……, ……

Figure 1.C.5: Template for evaluation committees to track choice of winning ideas



Figure 1.C.6: Sample display of winning suggestions on the floor's notice board

TABLE 1.C.2. TRAINING CURRICULUM

Training type	Training content
Motivational training for supervisors (1 hour, 2 groups)	• Details about new suggestion system • Introduction of problem-solving culture • Definition of roles and responsibilities of supervisors – How to maintain good relationships with workers and be a good leader? – How to encourage workers to submit ideas (biweekly huddles and on-the-job encouragement)? – How to create an effective and friendly suggestion culture through positive and constructive feedback and recognition?
Technical training for supervisors (1 hour, 2 groups)	• Experience with line huddles • Overview of submitted suggestions • Technical advice – Difference between problem- and solution-oriented submissions – Examples of improvement areas – Typical questions from workers and corresponding answer options
Evaluation committee training (1 hour, 1 group)	• Details about new suggestion system • Success factors of suggestion system • Tasks of evaluation committee – Monitoring of supervisors – Evaluation of suggestions and alignment of three winners with managers – Reward of prizes and display of names – Development of implementation plan

TABLE 1.C.3. OVERVIEW OF HUDDLES

Line code	Floor	Treat-ment	SMS 1	Date 1st	SMS 2	Date 2nd	Date 3rd	Total	≥ 1	Average across all huddles: Attend. rate	Number workers	Workers per line	Duration in min.	Expect. ideas	Number tasks
1304001	1	1	13.01.2020	03.02.2020	08.02.2020			1	1	100%	25	25	15,0	0	4
1304002	1	1	13.01.2020	14.01.2020	08.02.2020	02.02.2020		2	1	56%	14	25	17,5	1	4
1304003	1	1	13.01.2020	14.01.2020	08.02.2020	03.02.2020		2	1	73%	20	28	15,0	1	4
1304006	1	0				26.02.2020		1	1	100%	29	29	10,0		4
1304008	1	1	13.01.2020	14.01.2020	08.02.2020	03.02.2020		2	1	93%	26	28	15,0	3	4
1304017	2	1	13.01.2020	21.01.2020	08.02.2020	08.02.2020		2	1	64%	35	52	10,0	11	4
1304020	2	1	13.01.2020		08.02.2020	10.02.2020		1	1	43%	25	58		12	4
1304021	2	0				11.02.2020		1	1	51%	20	39	10,0	8	4
1304022	2	1	13.01.2020	21.01.2020	08.02.2020	10.02.2020		2	1	77%	33	43	10,0	10	4
1304024	2	1	13.01.2020	14.01.2020	08.02.2020			1	1	75%	30	40	10,0		4
1304034	3	1	13.01.2020	14.01.2020	08.02.2020	11.02.2020	26.02.2020	3	1	99%	28	28	5,0	2	1
1304035	3	1	13.01.2020	14.01.2020	08.02.2020	11.02.2020	26.02.2020	3	1	100%	23	22	5,0	2	1
1304037	3	1	13.01.2020	14.01.2020	08.02.2020	11.02.2020	27.02.2020	3	1	88%	22	25	11,0	2	1
1304038	3	1	13.01.2020	14.01.2020	08.02.2020	11.02.2020	27.02.2020	3	1	100%	26	26	4,0	1	1
1304049	4	1	13.01.2020		08.02.2020			0	0						
1304051	4	1	13.01.2020	15.01.2020	08.02.2020			1	1	100%	27	27	5,0	80	4
1304053	4	1	13.01.2020	14.01.2020	08.02.2020	27.02.2020		2	1	95%	26	27	5,5	30	2
1304054	4	1	13.01.2020	19.01.2020	08.02.2020	11.02.2020	27.02.2020	3	1	95%	29	31	8,7	37	4
1304065	5	1	13.01.2020	14.01.2020	08.02.2020			1	1	100%	46	46	10,0	6	4
1304067	5	1	13.01.2020	14.01.2020	08.02.2020	12.02.2020		2	1	56%	23	40	10,0	6	2
1304068	5	1	13.01.2020	14.01.2020	08.02.2020	16.02.2020		2	1	79%	31	38	10,0	2	0
1304070	5	1	13.01.2020	14.01.2020	08.02.2020			1	1	100%	47	47	7,0	3	1
1304072	5	0				16.02.2020		1	1	98%	41	42	10,0	5	0
1304093	6	1	13.01.2020	14.01.2020	08.02.2020	12.02.2020		2	1	90%	25	28	12,5	1	4
1304094	6	1	13.01.2020	14.01.2020	08.02.2020	26.02.2020		2	1	100%	24	24	10,0		4
1304096	6	1	13.01.2020	15.01.2020	08.02.2020	24.02.2020		2	1	100%	29	29	10,0	2	1
1304098	6	0				12.02.2020		1	1	100%	27	27	12,0	0	3
1304099	6	1	13.01.2020	14.01.2020	08.02.2020			1	1			23	10,0		4
1304102	6	1	13.01.2020	14.01.2020	08.02.2020			1	1	100%	23	23	1,2		4
Mean										**86%**	**28**	**33**	**9,6**	**10**	**3**
Nb. lines		**29**		23		21	5								
% of all lines				79%		72%	17%								
Nb. treated lines		**25**		23		17	5								
% of treated lines				92%		68%	20%								

Notes: The data in Table 1.C.3 is based on the huddle forms filled out by each supervisor at the end of a line huddle. "Treatment" is an indicator equal to one if the respective line was randomly assigned to treatment. "Total" is the total number of huddles conducted per line during the intervention period, and "≥ 1" is an indicator for at least one huddle. "Attend. rate" is the average attendance rate across all huddles, which is the number of workers attended ("Nb workers") divided by the number of workers per line ("workers per line"). The latter is taken from the baseline data. "Duration in min." shows the average duration of one line huddle per line. "Expect. ideas" is the average number of ideas a supervisor expects to be submitted by workers on his line. An empty cell means that the supervisor did not fill out this information. "Nb tasks" is the average number of tasks a supervisor has undertaken per line huddle. Supervisors could select up to four tasks: Reminding workers of the suggestion box, motivating workers to submit ideas, explaining how to fill out a suggestion paper, and answering questions of workers.

TABLE 1.C.4. OVERVIEW OF WINNING SUGGESTIONS

Line	Prize	Date of reward	Date of submission	Source	Category	Subcategory	Suggestion summary	Suggestion text (shortened)	Type	Total points
1304040	1	20.02.2020	26.12.2019	lunch	physical work env.	workwear or comfort	provide uniform/mask/id	Everyone should follow the rules and regulations in our factory. We should wear personal masks with our ID to be healthy and increase our output.	idea	16
1304023	1	20.02.2020	26.12.2019	lunch				We need to come to the office on time, avoid excessive talking and jealousy, and work together.	general observation	0
1304096	1	20.02.2020	08.01.2020	lunch	facilities/ convenience	transportation	provide transportation	Due to unavailability of life and transport facility, we can't make it on time to work. We should get both the facilities considering our timely arrival at work.	idea	16
1304055	1	20.02.2020	08.01.2020	lunch	health and safety	training	provide training/ training center	Skilled manpower is required to speed up the work, increase effectiveness and shipment on time. Company should provide training for the best outcome.	idea	18
1304068	1	20.02.2020	15.01.2020	lunch	skills	skill assignments	assign right people in right position	Workers should be assigned tp work based on their skills.	idea	21
1304017	1	22.03.2020	30.01.2020	lunch	psychological/ behavioral	workers behaviour	come to office on time/be attentive	We should come to office on time, be attentive, ensure work quality and not be absent without reason. We should also wear ID cards and masks to ensure safety.	idea	22
1304038	1	22.03.2020	05.02.2020	lunch	efficiency	production targets	less wastage of materials	We should reduce wastage while working.	problem/ behaviour	0
1304097	1	22.03.2020	05.02.2020	lunch	machine	functioning	machine malfunctioning	It causes problems if the needle breaks. We should be more attentive and manage time while working.	problem/ behaviour	0
1304071	1	22.03.2020	06.02.2020	box				Everyone should follow company rules, work attentively, reduce higher absenteeism, do the work they are good at and work together to enhance production.	general observation	0
1304021	1	22.03.2020	27.01.2020	box	psychological/ behavioral	workers behaviour	be attentive	We should work attentively to increase efficiency and ensure security. We should also reduce wastage while working.	idea	22
1304001	1	22.03.2020	12.02.2020	lunch				Extra clothes should be provided to reduce defects in finished goods.	general observation	0
1304040	2	20.02.2020	26.12.2019	lunch				I don't have any problem since the mechanics are provided when required.	no problem	0
1304018	2	20.02.2020	01.01.2020	lunch	psychological/ behavioral	workers behaviour	come to office on time	We should come to work on time and work properly to improve quality.	idea	22
1304066	2	20.02.2020	01.01.2020	lunch	absenteeism	attendance bonus	provide attendance bonus	Our attendance bonus should be paid.	idea	18
1304001	2	20.02.2020	25.12.2019	box	skills	operator skills	lack of skilled workers	Production and quality are getting hampered by the lack of skilled workers, hence there is a delay in product delivery.	problem/ behaviour	0
1304055	2	20.02.2020	02.01.2020	box	absenteeism	leave	provide/ grant leave	We should be provided with our leaves and attendance bonus.	idea	23
1304100	2	20.02.2020	15.01.2020	lunch	absenteeism	leave	provide/ grant leave	Our leaves need to be granted to reduce absent rate in factory.	idea	23
1304050	2	22.03.2020	23.01.2020	lunch	manpower	lack of general manpower	provide manpower	Sufficient manpower should be given in each line and they should work attentively and efficiently to fulfill their targets.	idea	18
1304039	2	22.03.2020	23.01.2020	lunch	psychological/ behavioral	workload and pressure	work in all lines/no fixed line	We don't have any fixed line to work at, hence the production rate and quality is low.	problem/ behaviour	0
1304019	2	22.03.2020	23.01.2020	lunch	psychological/ behavioral	workers behaviour	be attentive	We should work attentively to have a quality output. Everyone should work together and the supervisors should motivate the workers.	idea	22
1304070	2	22.03.2020	30.01.2020	lunch	absenteeism	attendance bonus	provide attendance bonus	We should be provided with attendance bonus.	idea	18
1304100	2	22.03.2020	30.01.2020	lunch	absenteeism	leave	provide leave if someone is ill	We should get leaves when we are ill so that we can work in good health and be motivated towards work.	idea	22
1304002	2	22.03.2020	03.02.2020	box	psychological/ behavioral	workers behaviour	be attentive	We need to work attentively to avoid alters and defects. The critical works should be closely monitored.	idea	22
1304036	3	20.02.2020	01.01.2020	lunch				We should all work together to imrove the work environment.	general observation	0
1304068	3	20.02.2020	01.01.2020	lunch	absenteeism	leave	provide/ grant leave	Our leaves should be granted so that we don't remain absent.	idea	23
1304004	3	20.02.2020	25.12.2019	box	absenteeism	attendance bonus	provide attendance bonus	Percentage of absenteeism will decrease and productivity will increase if we get attendance bonus and leaves.	idea	18
1304022	3	20.02.2020	15.01.2020	lunch	facilities/ convenience	place for prayer	no place for prayer	We don't have a perfect place for prayers. Work environment can be improved if supervisors show us sympathy and help us.	problem/ behaviour	0
1304038	3	20.02.2020	15.01.2020	lunch	physical work env.	air quality	bad odor from toilet	Bad odor comes from toilet near the input line.	problem/ behaviour	0
1304099	3	20.02.2020	15.01.2020	lunch	absenteeism	leave	provide/ grant leave	Absenteeism will be reduced if we are granted our leaves timely.	idea	23
1304053	3	22.03.2020	23.01.2020	lunch	machine	availability	availability of machines	We face problems due to machine malfuntioning, line changing and not granting any leaves.	problem/ behaviour	0
1304097	3	22.03.2020	23.01.2020	lunch	absenteeism	attendance bonus	provide attendance bonus	We will be benefited if attendance bonus and sufficient helpers are provided. We can give full effort if we are not scolded.	idea	18
1304069	3	22.03.2020	30.01.2020	lunch	motion waiting time	movement of workers	less movement of workers	Excess movement of workers need to be reduced.	problem/ behaviour	0
1304040	3	22.03.2020	25.01.2020	box	physical work env.	cleanliness	use machine cover while dusting	Machines should be covered while dusting and cleaning should be done after working hours to avoid dirt.	idea	21
1304021	3	22.03.2020	27.01.2020	box	absenteeism	leave	provide/ grant leave	Efficiency will increase if we get our leaves, work attentively and keep safety measures.	idea	23
1304001	3	22.03.2020	12.02.2020	lunch	absenteeism	leave	provide/ grant leave	Rate of absenteeism will decrease if leaves are granted. We should come to the office on time and supervisors should solve our problems.	idea	23

Notes: Table 1.C.4 gives an overview of all winning suggestions which have been selected by the evaluation committees over the course of the interventions. A prize of one is rewarded with 500Tk., a prize of two with 300Tk. and a prize of three with 100Tk. Suggestions can be from the suggestion box (box) or from the active managerial intervention (lunch). Type and total points are assessments by an industry expert, who is independent from the factory.

1.D Sixth-step randomisation approach into lunch treatment

The randomisation of 1.599 workers into lunch and non-lunch groups was implemented at the pair level (dyads of workers) to have a similar number of friend and non-friend dyads in both groups. Dyads in the control group were never paired in reality. I carried out the randomisation in six steps as described below.

i. Split of sample into workers with and without friends: To begin with, I split the sample into workers with friends on the same production floor (sample one) and without friends (sample two). The first sample (sample one) consists of 1439 workers with friends on the same floor. 714 have one friend and 725 have two friends on the same floor. The second sample (sample two) consists of 160 workers. 51 of them have friends on a different floor (group one), 80 have no friends but are mentioned by someone else (group 2), and 39 have no friends and are also not mentioned by anyone else on the same floor. The numbers are also displayed in Table 1.D.5. The reason for these categorizations is that workers with friends might have different characteristics or behaviours than workers without. By excluding the latter, I can test whether results still hold, in which case they would not behave differently.

TABLE 1.D.5. OVERVIEW OF TWO RANDOMISATION SAMPLES

Total number of workers	**1599**
SAMPLE 1	**1439**
- with one friend on the same floor	714
- with two friends on the same floor	725
SAMPLE 2	**160**
- group 1: with friends on a different floor	41
- group 2: without friends (mentioned by someone else on the same floor)	80
- group 3: without friends (not mentioned by someone else on the same floor)	39

ii. Generation of dyads: As a next step, I generate pairs of workers. This step is first done for workers with friends (sample one), and then for workers without friends (sample two). There are two instances, in which workers are moved from sample two into sample one, or the other way round. On the one hand, some workers in sample two, which do not mention any friends, but are mentioned by someone else as a friend, could be 'drawn' into sample one as I assume that friendships are symmetric. On the other hand, some workers in sample one could become part of sample two if they are not able to be paired, which is depending on the density of the network.

Dyads of friends: In order to generate as many dyads of friends as possible in a scarce network, I pair workers in a particular order depending on the number of friends and the number of times a worker is mentioned by someone else as

a friend. I first pair workers with only one friend and then workers with two friends. Within both groups, I first pair workers who have been mentioned least by someone else as a friend and lastly pair workers who have been mentioned most by someone else, i.e. I start with the most unpopular workers (zero times mentioned by someone else) and end with the most popular workers (eleven times mentioned by someone else). The distribution of the number of times being mentioned as a friend is shown in Figure 1.D.7. As the number of possible matches for unpopular workers is lower than for popular workers, this approach allows us to come up with the highest number of possible matches in a scarce network. Workers who cannot be paired are added to sample two (group 4: 'leftovers').

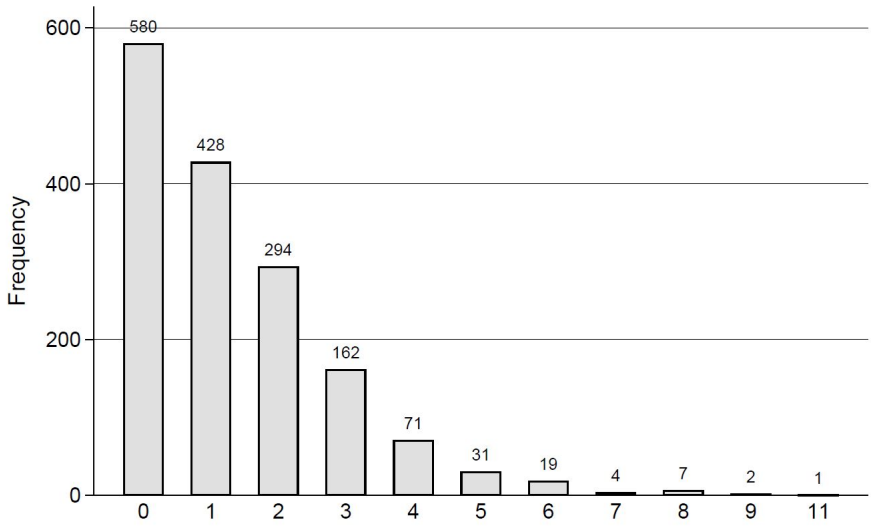


Figure 1.D.7: Frequency of workers by number of times being mentioned as friend

Dyads of non-friends: I generate as many dyads of non-friends as possible for workers in sample two, i.e. four sub-groups of workers: Workers with friends on different floors (group 1), workers without friends but mentioned by someone else on the same floor (group 2), workers without friends but not mentioned by anyone on the same floor (group 3), and workers who could not be paired in the previous step (group 4). The matching process is randomly performed by floor and by sub-group. Very few workers are not able to be paired at all.

iii. Randomisation into lunch and non-lunch group: I now randomise the dyads of friends/non-friends into treatment and control group by sample such that sample one and sample two consists of a balanced number of workers in both groups. I hereby take into account whether a dyad consists of at least one worker who had been asked an extended survey question or not. This ensures to have a balanced number of extended survey workers in both groups for each line. Hence, I first generate dyads with at least one extended survey worker stratified by sample and by line. All uneven remaining dyads are randomly assigned by sample only. I then generate dyads with no extended survey workers stratified by sample and by floor. All uneven remaining dyads are completely randomly assigned.

iv. Re-balancing: I re-balance treatment assignments to achieve balance on the maximum experience index per dyad across treatment groups and on the percentage of treatment dyads across lines. To do so, I repeat above mentioned steps ii. and iii. 100 times and randomly select one assignment which fulfills the following two criteria:

- The p-value of the differences in maximum experience indices per dyad should be larger than 0.2. I select the maximum experience index per dyad as I expect the performance of a pair to be determined by the highest experience index, and not by the average.
- The top decile of the difference in the number of treatment and control dyads per line (as a percentage of all dyads per line) of sample 1 should be < 0.45 and of sample 2 < 0.75. Effects might be different if a supervisor is exposed to fewer or more lunch workers. It would be interesting to exploit that difference, but I do not have enough power to do that.

v. Re-pairing of friend dyads: As a penultimate step, I randomly select 50% of the dyads (out of the friend dyads from sample one) to be not seated with a friend. These dyads are taken apart and workers are re-paired with a non-friend on the same floor. This step is done to be able to compare friend and non-friend dyads within sample one, i.e. within workers who have friends. The final number of workers per treatment assignment is shown in Table 1.D.6.

TABLE 1.D.6. TREATMENT ASSIGNMENTS

SAMPLE 1	**1244**
Workers in treatment group (friends and non-friend dyads)	**624 (312 pairs)**
- hereof workers in friend dyads	310 (155 pairs)
- hereof workers in non-friend dyads	314 (157 pairs)
Workers in control group	**624 (312 pairs)**
SAMPLE 2	**340**
Workers in treatment group (non-friend dyads only)	**170 (85 pairs)**
– group 1: friends on different floor	16
– group 2: without friends (mentioned by someone else on the same floor)	8
– group 3: without friends (not mentioned by someone else on the same floor)	18
– group 4: leftovers from sample one	128
Workers in control group	**170 (85 pairs)**
– group 1: friends on different floor	20
– group 2: without friends (mentioned by someone else on the same floor)	8
– group 3: without friends (not mentioned by someone else on the same floor)	18
– group 4: leftovers from sample one	126
Not in any sample / unpaired	**11**

vi. Change of seating order: After lunch rounds one and two, I randomly select 1/3 of all pairs assigned to lunch – 1/3 of the friend and non-friend pairs in sample one and 1/3 of the non-friend pairs in sample two – to be seated opposite a new non-friend, as shown in Table 1.D.8. I first dropped all workers who have left the factory, and indicate whether a worker's partner has left (see Table 1.D.7).

TABLE 1.D.7. NUMBER OF EXITS BEFORE CHANGE OF SEATING ORDER

Workers in treatment group	795
- hereof left	104
Remaining	**691**
Workers in control group	794
- hereof left	82
Remaining	**712**

I then balanced on an indicator equal to one if a worker's seating partner had left the factory by the time the change of the seating partners was arranged, i.e. I randomly chose 1/3 of workers whose partner had left, and 1/3 of workers whose partner had not left. Next, I re-paired all those workers within sample one and within sample two with a non-friend by floor. Finally, I randomly chose a seating order for the dyads. Workers whose partner had left (but have not been re-paired) were randomly seated at the edges of the tables, because I could not leave empty spaces at a table. The number of workers who had new seating partners per sample are shown in Table 1.D.8.

TABLE 1.D.8. CHANGE OF SEATING ORDER

SAMPLE 1	549
Seating change	182
- hereof friend dyads	90
- hereof non-friend dyads	92
Unchanged	367
- hereof friend dyads	181
- hereof non-friend dyads	186
SAMPLE 2	141
Seating change	49
Unchanged	92

1.E Components of main outcomes

TABLE 1.E.9. PRIMARY OUTCOME VARIABLES

PRIMARY OUTCOMES RELATED TO FIRM PERFORMANCE		
WORKFORCE INDEX		
absenteeism	Ratio: number of workers absent / number of total workers (present + absent)	daily clock data
punctuality	Minutes worker arrives before production start date	daily clock data
working hours	Working hours in minutes	daily clock data
Other workforce-related variables (not included in index)		
absenteeism on treatment day	Ratio: number of workers absent on treatment day / number of total workers on treatment day	daily clock data
exits	Ratio: number of workers left / number of total workers	monthly salary data
PRODUCTION INDEX		
efficiency	Ratio: output minutes (smv*daily output) / number of workers	weekly production data
alteration rate	Ratio: number of alterations / number of pieces checked	weekly production data
reject rate	Ratio: number of rejects / number of pieces checked	weekly production data
PRIMARY OUTCOMES RELATED TO WORKER WELL-BEING		
JOB SATISFACTION INDEX (WORKERS)		
Job satisfaction subindex		
own job satisfaction	Own job satisfaction (scale 1-4)	survey data
others' job satisfaction	Co-workers' job satisfaction (scale 1-4)	survey data
career opportunities	Satisfaction with career opportunities at work (scale 1-4)	survey data
job training & support	Satisfaction with level of job training and support (scale 1-4)	survey data
Turnover intentions subindex		
intent to quit	Number of times thinking about quitting your job in the last three weeks	survey data
intent to stay	Number of days you would like to continue working in this factory	survey data
Job attitudes subindex		
own attitude about job	Own attitude to do the best work you can, even if it sometimes does interfere with the rest of my life (scale 1-3)	survey data
others' attiude about job	Other operators' attitude to do the best work they can, even if it sometimes does interfere with the rest of their life (scale 1-3)	survey data
sense of punctuality	Minutes worker should arrive before work start	survey data

Continued on next page

TABLE 1.E.9 – continued from previous page

Job satisfaction of supervisors (not included in index)		
Job satisfaction subindex		
own job satisfaction	Satisfaction with own job (scale 1-4)	survey data
operators' job satisfaction	Satisfaction of operators with their job (scale 1-4)	survey data
happiness at work	Happiness while at work (scale 1-10, with ten being very happy, five being sometimes happy and sometimes unhappy and 1 being very unhappy)	survey data
Turnover intentions subindex		
intent to quit	Number of times thinking about quitting your job in the last three weeks	survey data
intent to stay	Number of days you would to continue working in this factory	survey data
MENTAL HEALTH INDEX		
Feelings at work subindex		
feeling angry	Number of times you feel angry, hurt, stressed, or frustrated because of a situation that arose at work	survey data
person involved	Number of persons involved in causing the mood	survey data
feeling happy	Number of times you feel happy, cared-for, or good because of a situation that arose at work	survey data
person involved	Number of persons involved in causing the mood	survey data
General anxiety disorder subindex		
nervous	Feeling nervous, anxious, or on edge	survey data
sleepless	Not being able to sleep or control worrying	survey data
worrying	Worrying too much about different things	survey data
trouble relaxing	Trouble relaxing	survey data
restless	Being so restless that it is hard to sit still	survey data
annoyed	Becoming easily annoyed or irritable	survey data
afraid	Feeling afraid, as if something awful might happen	survey data
Workplace harassment subindex		
frequency of abusive language	Number of days other operators on your line needed to put up with shouting or abusive language at work over the last two weeks (scale 1-4)	survey data
frequency of uncomfortable situations	Number of days other operators on your line had to experience situations at work that made them feel uncomfortable, for example persistent staring, flirting, grabbing their hand or other parts of the body, etc. (scale 1-4)	survey data
touched by primary supervisor	Dummy: Primary supervisor every physically touched you in aggressive way	survey data
touched by any supervisor or manager	Dummy: Any supervisor or manager ever talked to you or touched in a sexual way	survey data
PRIMARY OUTCOMES RELATED TO INNOVATION		
quantity	Indicator equal to one for any submitted suggestion and equal to zero for no submitted suggestion per round	intervention output

Continued on next page

TABLE 1.E.9 – continued from previous page

type	Type of suggestion from one to six: 1 incoherent, 2 general observation/no problem, 3 poorly communicated problem, 4 well communicated problem, 5 poorly communicated idea, 6 well communicated idea	intervention output
quality of ideas	Quality index for ideas only (type 5 and 6) consisting of six components on a scale from one to five each (apart from component four which only has a scale of one to two); index constructed using five clusters of <20, 20, 21-22, 23, and 24-27 points: 1. **Implementation level:** At which level can the idea be best implemented? (one workplace, specific workplaces, line level, floor level, or factory level) 2. **Impact on worker well-being:** How large would be the expected benefit for affected workers? (very low to very high) 3. **Impact on efficiency:** How large would be the expected effect on productivity? (very low to very high) 4. **Level of effects:** Are productivity effects direct (e.g. repaired machine) or indirect through improved motivation of workers (e.g. prayer place)? 5. **Implementation difficulty:** How fast and riskless is the suggestion to implement? (very long to very fast) 6. **Financial resources:** How much money is needed for implementation? (very substantial over BDT 50k to none)	intervention output
quality	Quality index of suggestions ranging from zero to ten, which is an additive combination of the type and - for type five and six (ideas) - the quality index of ideas, non-submissions are included with a value of zero	intervention output
quality (if submitted)	Same as quality index above, but non-submissions are excluded, i.e. ranging from one to ten	intervention output
worker components	Same as quality index above, but quality index of ideas only measures impact on worker outcomes (components three and four)	intervention output
firm components	Same as quality index above, but quality index of ideas only measures impact on firm outcomes (components five to eight)	intervention output
content	Indicator equal to one if submitted suggestion is classified as respective category; twelve different indicators for the following categories: Supervisor skills, leave, attendance bonus, work environment (behavioral), work environment (facilities), manpower, machine, salary, motion/waiting time, efficiency, health and safety, and production layout	intervention output
variety	Number of submitted suggestions per round which have not been mentioned by worker before	intervention output

TABLE 1.E.10. VARIABLES TO MEASURE HETEROGENEOUS EFFECTS

worker's experience index	Anderson index including number of factories worked at, grade, tenure in factory in years, tenure in industry in years, years of education	survey data
processes	Number of sewing processes a worker can perform	skills data
worker's gender	Indicator equal to one for female worker	survey data
new seating order	Indicator equal to one for workers who were seated next to a new partner in lunch round three	randomisation data
friend as seating partner	Indicator equal to one for workers who were seated opposite a friend	randomisation data
friendship intensity	Number of topics worker discusses with friend and number of locations	survey data
number of friends	Number of friends as seating partners (left, right, opposite, diagonally left, diagonally right)	intervention output
cooperativeness	Similarity measure of suggestions from seating partners	intervention output
supervisor's experience index	Anderson index including number of factories worked at, tenure in factory in years, tenure in industry in years, years of education	survey data
supervisor's gender	Indicator equal to one for female worker	survey data
perceived efficacy supervisors	Beliefs about winning and implementation, and assessment of workers	survey data

Category	Sub-category	Suggestion paper	Idea	Problem
Supervisor skills	skill assignments	mismanagement in work allocation		✓
		assign right people in right position	✓	
	supervisor gender	female supervisors wanted		✓
		female supervisors not wanted		✓
		need male supervisor	✓	
	supervisor skills	misbehaved/scolded by supervisors		✓
		inconsiderate/less supportive supervisors		✓
		no guidance in newly assigned work		✓
		work done forcefully		✓
		appoint better supervisor	✓	
		avoid scolding/misbehaviour	✓	
		need more support from supervisors	✓	
		need raw material on time	✓	
		provide educated supervisor	✓	
		provide good behaviour	✓	
Attendance bonus	attendance bonus	no attendance bonus		✓
		provide attendance bonus	✓	
Leave	leave	too high absenteeism		✓
		understand reasons for leave/leave not granted		✓
		walk around a lot to request leave		✓
		provide/grant leave	✓	
		provide leave if someone is ill	✓	
Work environment behavioral/ psychological	workers behaviour	unnecessary chaos/chatting on floor		✓
		avoid harassment	✓	
		avoid unnecessary talking	✓	
		be attentive	✓	
		come to office on time	✓	
		follow rules and regulations	✓	
	workload and pressure	partial behaviour in workplace		✓
		excessive workload		✓
		work in all lines/no fixed line		✓
		avoid partiality	✓	
		assign working on fixed line	✓	
		leave on time	✓	
		reduce work pressure	✓	
Work environment facilities/ convenience	lift	no availability of lift		✓
		technical difficulties of lift		✓
	toilet	not allowed to go to toilet		✓
	place for prayer	not allowed to pray		✓
		no place for prayer		✓
		provide prayer place for girls	✓	
		provide prayer place/mosque	✓	
	punch machine	no punch machine		✓
		provide punch machine	✓	
	transportation	no transportation		✓
		provide transportation	✓	
	road condition	clear road from vehicles to cross safely	✓	
		provide foot overbridge	✓	
	water facility	need heater for warm water	✓	
		need water during lunch	✓	
	lunch	provide lunch from factory	✓	
		allow to bring food on floor	✓	
	yearbook	publish annual/yearbook	✓	

Continued on next page

Figure 1.E.8: Categorizations of suggestions

Figure 1.E.8 – continued from previous page

Category	Sub-category	Suggestion paper	Idea	Problem
Work environment workplace	air quality	bad odor from toilet		✓
		affected by polluting air		✓
	light	not enough light		✓
	temperature	less/no fans		✓
		cold temperature		✓
		wind from outside		✓
		hot temperature		✓
		provide fans	✓	
	cleanliness	dirty working place/toilet		✓
		assure cleanliness of workplace	✓	
		avoid spitting/littering	✓	
		provide soap	✓	
		use machine cover while dusting	✓	
	damages	broken chair/table		✓
		fix broken chair	✓	
		repair damaged floor tiles	✓	
	noise	place sounbox in middle	✓	
		play national anthem in the morning	✓	
		no music after lunch		✓
	workwear or comfort	no shoes allowed		✓
		provide chair/table	✓	
		provide cushion	✓	
		provide shoe boxes	✓	
		provide uniform/mask/id	✓	
	box	provide boxes	✓	
Manpower	replacement	no quick replacement for operators		✓
	abundance of manpower	excessive staffs in workplace		✓
		contractual workers not wanted		✓
		reduce excessive staffs	✓	
		reduce excessive workers	✓	
	helpers	no helper available		✓
		provide more helpers	✓	
	lack of general manpower	lack of skilled workers		✓
		need more operators	✓	
		appoint better/skilled workers	✓	
		provide manpower	✓	
Machine	availability	availability of machines		✓
		replace old machines	✓	
	functioning	machine malfunctioning		✓
		delay in solving machine problems		✓
		machines are not clean		✓
		need to clean machines	✓	
		provide eye/needle guard	✓	
		provide good mechanic	✓	
		provide magnet to find needle	✓	
		quality checks of machines	✓	

Continued on next page

Figure 1.E.8 – continued from previous page

Category	Sub-category	Suggestion paper	Idea	Problem
Salary	base salary	insufficient salary		✓
		need salary on time	✓	
		provide more salary	✓	
		provide salary in hand	✓	
	bonus salary	not enough night bill/overtime		✓
		provide benefits/increment/production bonus	✓	
	fairness	partiality in benefits/increment		✓
		salary discrimination		✓
		need promotion for helpers	✓	
		provide fair salary based on hierarchial level	✓	
		reduce discrimination	✓	
Motion/ waiting time	non-productive time	false fire alarm		✓
		waiting for partner to come		✓
	movement of workers	less movement across lines and processes		✓
	location of canteen	canteen too far away		✓
		need canteen at nearest location	✓	
Efficiency	labelling	bundle cut should reach output table		✓
		maintain serial after bundle cut	✓	
		provide correct body stickers	✓	
	production targets	lack of extra bobbin		✓
		less wastage of materials		✓
		need quicker process in production	✓	
		measure production rate	✓	
		need more target	✓	
	quality check	change cutter	✓	
		need to check quality	✓	
Health and safety	health tools/equipment	not wearing mask		✓
	pregnancy	many movements for pregnant woman		✓
	safety measures	no CCTV		✓
		provide cctv camera	✓	
		form a safety committee	✓	
	medical care	longer availability of doctors from 8am till night	✓	
		provide eye test facility	✓	
	training	provide training/training center	✓	
Production layout	error	change layout		✓
		shipment problem		✓

Notes: Figure 1.E.8 shows the categorizations of all submitted suggestions into category, subcategory and suggestion paper by an industry expert. A tick in the idea column indicates that respective suggestions were assessed as the type "idea". A tick in the problem column indicates that respective suggestions were assessed as the type "problem".

	Survey components	
	✓ included	
	× excluded	
Survey questions	**Operators**	**Supervisors**
Index: Engagement and support	✓	✓
Subindex: Idea generation	✓	✓
Frequency thinking about problems	✓	✓
Frequency thinking about solutions	✓	✓
Thinking about ideas over lunch	✓	×
Frequency of discussing ideas with people outside of the factory	✓	×
Feeling responsible for changes on the production line	✓	✓
Subindex: Leaderhip efficacy	✓	✓
Ability of typical supervisor to encourage operators	✓	✓
Ability of own supervisor/yourself to encourage operators	✓	✓
Difference in ability to encourage operators between typical and own supervisor	✓	✓
Overall ability of typical supervisor	✓	✓
Overall ability of own/yourself supervisor	✓	✓
Difference in overall ability between typical and own supervisor	✓	✓
Subindex: Submission of suggestions	✓	✓
Ever submitted an idea into the suggestion box	✓ Endline, all	×
Ever told an idea to someone	✓ Endline, all	×
Person to whom idea was told is an operator	✓ Endline, all	×
Estimated number of suggestions	×	✓ Endline
Frequency of knowing content of suggestions	×	✓ Endline
Knowledge about submitted topics on your line	×	✓ Endline
Own ideas communicated downwards	×	✓ Endline
Own ideas communicated upwards	×	✓ Endline
Index: Perceived constraints	✓	✓
Subindex: Constraints operators	✓	✓
Time as a constraint	✓	✓
Knowledge as a constraint	✓	✓
Incentives as a constraint	✓	✓
Decision-making authority as a constraint	✓	✓
Uncomfortable feeling as a constraint	✓	✓
Subindex: Constraints supervisors	×	✓
Time as a constraint	×	✓
Knowledge as a constraint	×	✓
Incentives as a constraint	×	✓
Decision-making authority as a constraint	×	✓
Uncomfortable feeling as a constraint	×	✓
Subindex: Ease of idea generation	✓	×
Easiness to think about problems	✓	×
Easiness to think about solutions	✓	×
Subindex: Ease of submission	✓	×
Feeling incentivized to submit idea	✓ Endline, all	×
Feeling comfortable to submit idea	✓ Endline, all	×

Continued on next page

Figure 1.E.9: Secondary outcome variables

Figure 1.E.9 – continued from previous page

Survey questions	Operators	Supervisors
Index: Perceived efficacy	✓	✓
Subindex: Belief about winning and implementing	✓	✓
Self-assessment of own submitted suggestions	✓	✗
Likelihood of winning	✓	✗
Likelihood of implementation of winners' idea	✓	✗
Likelihood of implementation of own idea	✓	✗
Likelihood of prize on own line	✗	✓
Likelihood of implementation by managers	✗	✓
Subindex: Assessment of workers	✗	✓
Motivation of typical worker in the factory	✗	✓
Motivation of typical worker on your line	✗	✓
Concentration of typical worker in the factory	✗	✓
Concentration of typical worker on your line	✗	✓
Ability to develop suggestions of typical worker in the factory	✗	✓
Ability to develop suggestions of typical worker on your line	✗	✓
Subindex: General perceptions	✗	✓
Workers as biggest beneficiary of suggestion box	✗	✓ Endline
Assessment of group meetings	✗	✓ Endline
No concerns such as workers complaining about supervisors	✗	✓ Endline
Submission of suggestions which workers would otherwise not tell anyone	✗	✓ Endline
Satisfaction with award of prizes	✗	✓ Endline
Index: Communication intensity	✓	✓
Subindex: General communication	✓	✓
Number of topics discussed with supervisor	✓	✗
Something discussed	✓	✗
Work-related grievances/problems discussed	✓	✓
Improvement of skills discussed	✓	✓
Production capabilities discussed	✓	✓
Work environment discussed	✓	✓
Frequency of discussions / number of workers talked to more than three times over last two weeks	✓	✓
Forms of discussions (individual meetings, group meetings, spontaneously on the line and/or during breaks)	✓	✓
Worker involved in initiating the discussion	✓	✓
Subindex: Raising suggestions	✓	✗
Problem raised to supervisor	✓	✗
Suggestion made to supervisor	✓	✗
Problem raised to manager	✓	✗
Suggestion made to manager	✓	✗
Subindex: Suggestion box communication	✓	✓
Frequency of talking about suggestion box with supervisor	✓ Endline, all	✗
Frequency of line group meetings about suggestion box	✓ Endline, all	✗
Number of workers talked about box over last two weeks	✗	✓ Endline
Number of workers talked about box more than three times over last two weeks	✗	✓ Endline
Forms of discussions (individual meetings, group meetings, spontaneously on the line and/or during breaks)	✗	✓ Endline
Worker involved in initiating the discussion	✗	✓ Endline
Content of discussion (organizational issues, content of suggestions, other)	✗	✓ Endline

Continued on next page

Figure 1.E.9 – continued from previous page

Survey questions	Operators	Supervisors
Subindex: Communication with managers	×	✓
Frequency of talking about work-related grievances/problems with manager	×	✓
Frequency of talking about skills with manager	×	✓
Frequency of talking about production capabilities with manager	×	✓
Frequency of talking about work environment with manager	×	✓
Forms of discussions (individual meetings, group meetings, spontaneously on the line and/or during breaks)	×	✓
Supervisor involved in initiating the discussion	×	✓
Encouragement to think about suggestions from managers	×	✓
Level of information about suggestion box from managers	×	✓ Endline
Index: Workplace culture	✓	✓
Subindex: Voice climate workers	✓	✓
Sharing of personal issues with superior	✓	×
Operators sharing issues with superior	✓	✓
Importance of sharing personal issues	✓	✓
Ability to be yourself at work	✓	×
Ability to openly state opinions	✓	✓
Fear of doing something wrong	✓	✓
Fear of admitting an error	✓	✓
Misreporting to supervisor	✓	✓
Degree of trust to workers (telling a secret)	✓	✓
Degree of trust to workers (lending money)	×	✓
Feeling valued if suggestion made	✓	×
Operator suggestion culture	×	✓
Subindex: Voice climate supervisors	×	✓
Ability to openly state opinions	×	✓
Fear of doing something wrong	×	✓
Fear of admitting an error	×	✓
Misreporting to managers	×	✓
Degree of trust to workers (telling a secret)	×	✓
Feeling valued if suggestion is made	×	✓
Supervisor suggestion culture	×	✓
Subindex: Caring culture	✓	×
Feeling listened to	✓	×
Feeling treated with respect	✓	×
Feeling appreciated & recognized	✓	×
Support from others	✓	×
Effort of supervisor to get to know worker	✓	×
Efforts to get to know each other	✓	×
Subindex: Mindset about authority	✓	✓
Conflict between workers and managers	✓	✓
Influence over workers	×	✓
Importance of group goal setting compared to supervisor goal setting	×	✓

1.F Supplementary figures

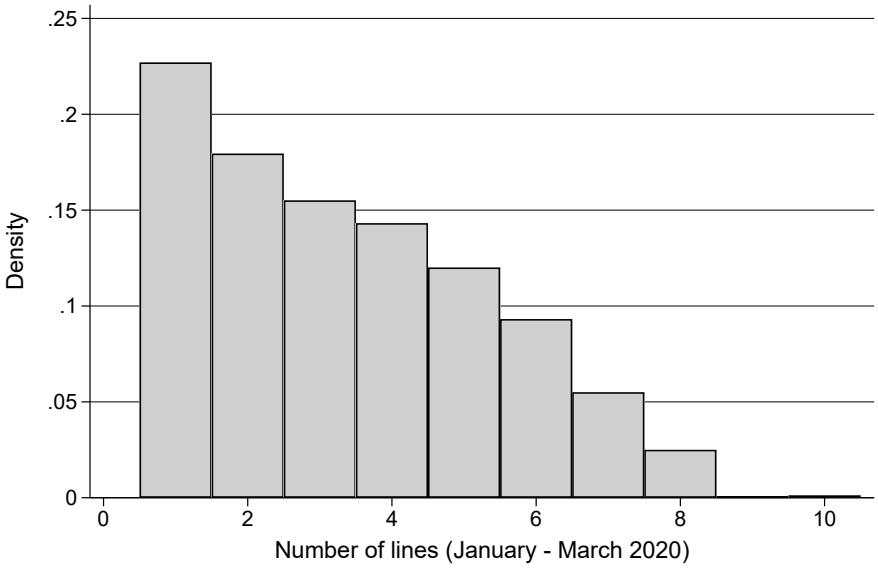


Figure 1.F.10: Distribution of total number of lines per worker

Notes: Graph 1.F.10 shows the distribution of the average number of lines each worker is assigned to over a period of three months, i.e. 91 days, between January 2020 and March 2020.

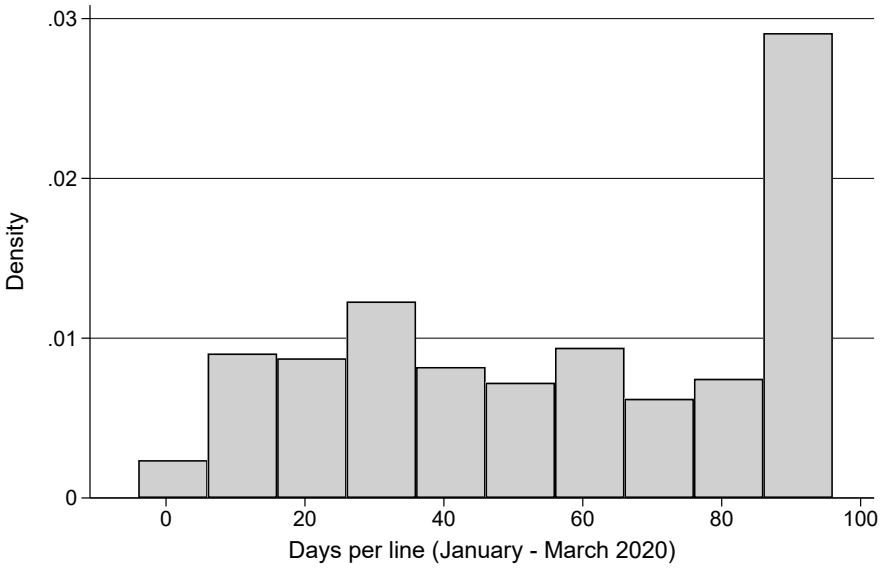


Figure 1.F.11: Distribution of maximum number of days per line per worker

Notes: Graph 1.F.11 shows the distribution of the average maximum number of days each worker is assigned to one line over a period of three months, i.e. 91 days, between January 2020 and March 2020.



Figure 1.F.12: Effects on primary outcome variables

Notes: Figure 1.F.12 shows treatment effects on primary outcome variables related to firm performance, represented by the workforce index and the production index, and worker well-being, represented by the job satisfaction index and the mental health index. Indices are based on Anderson (2008). Index components are shown in Table 1.E. Corresponding regression results are shown in Table 1.5. 95% confidence intervals are displayed.

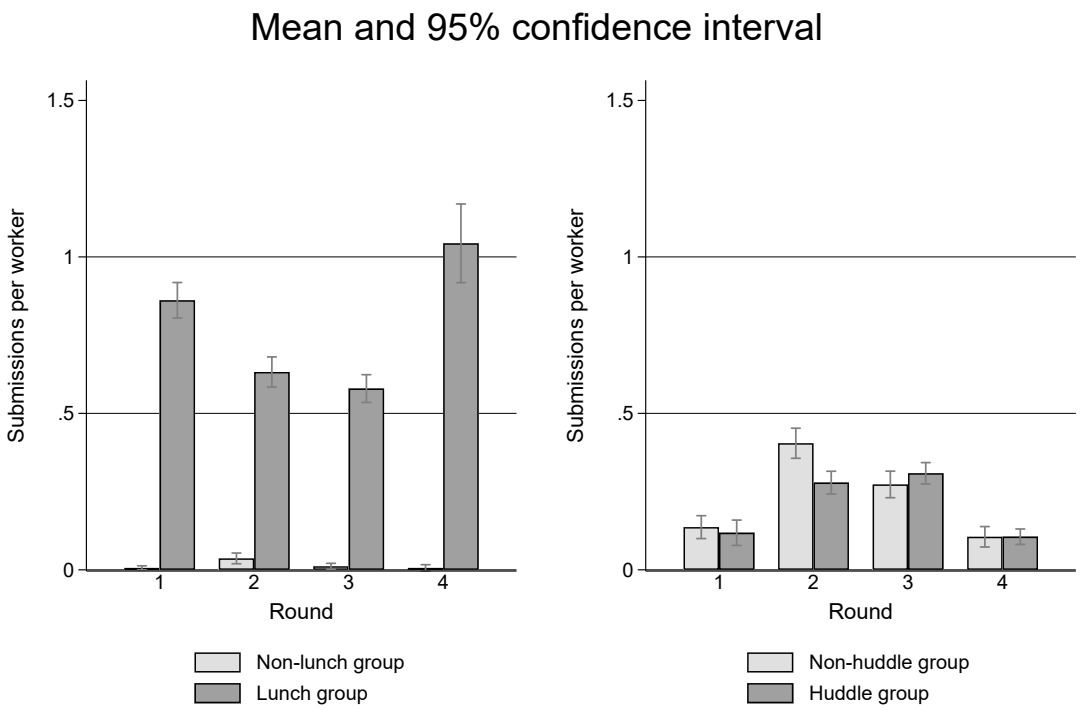


Figure 1.F.13: Average number of suggestions per round

Notes: The left graph of Figure 1.F.13 shows the mean and 95% confidence interval of the number of submitted suggestions by non-lunch and lunch workers per round. The fourth round restricts the data of lunch workers to those who were assigned to group one, in other words who received lunch in round four. The right graph shows the mean and 95% confidence interval by non-huddle and huddle workers. The data excludes all suggestions submitted prior to the date when the first huddle took place, i.e. 14th of January 2020.

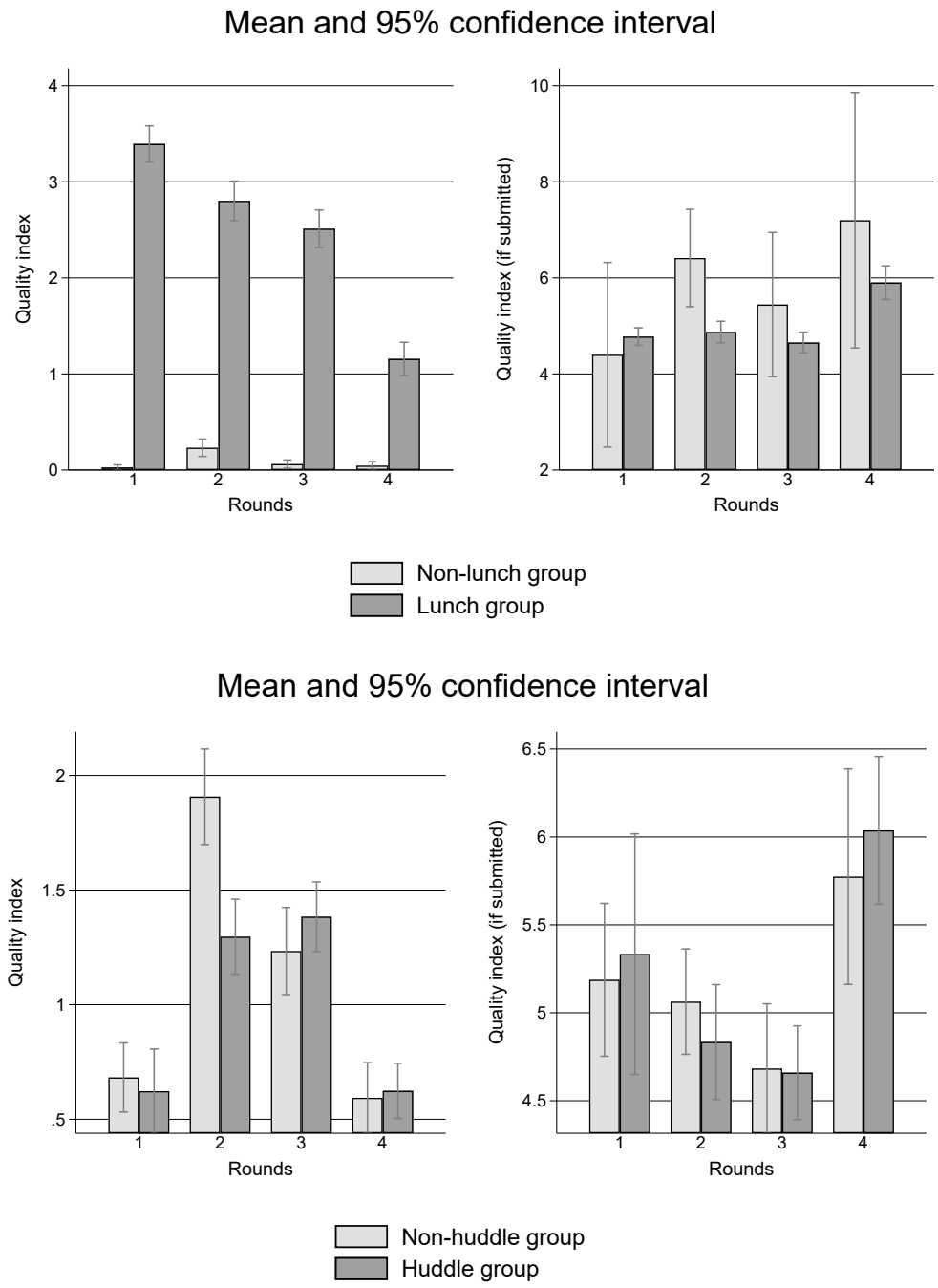


Figure 1.F.14: Quality of submitted suggestions per round

Notes: The upper graphs of Figure 1.F.14 shows the mean and 95% confidence intervall of the quality of suggestions by non-lunch and lunch group per round, whereas the lower graphs show it by non-huddle and huddle group per round. In both left graphs, the quality is set to zero if no suggestion was submitted in the respective round. In both right graphs, the data is restricted to submitted suggestions only.

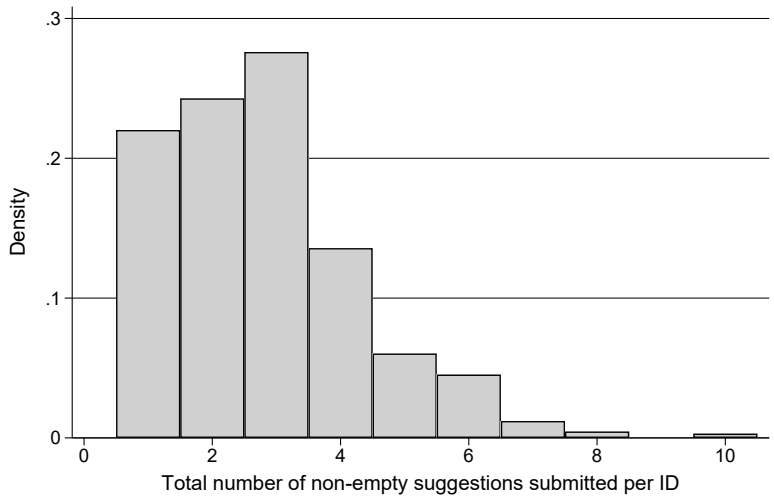


Figure 1.F.15: Total number of submissions per worker

Notes: Figure 1.F.15 shows the distribution of the total number of submissions for workers who submitted at least one suggestion.

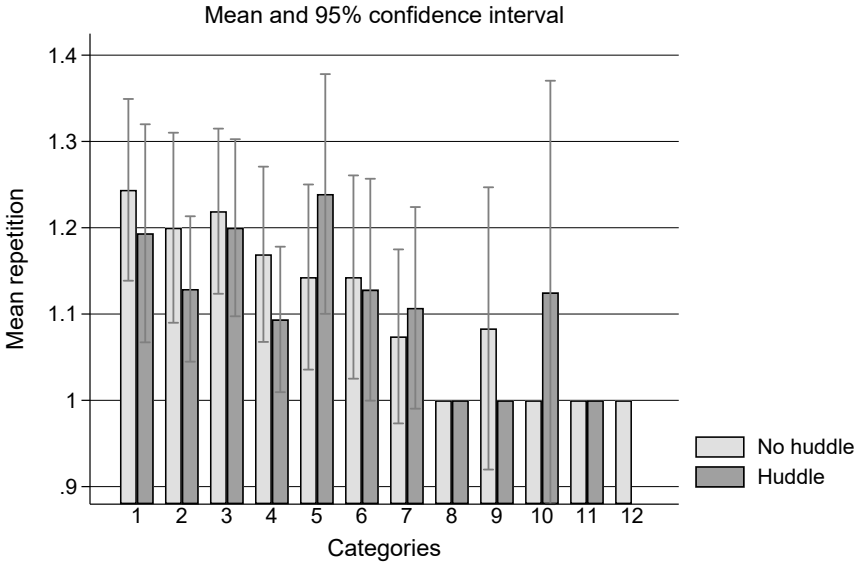


Figure 1.F.16: Repetition of categories by huddle treatment group

Notes: Figure 1.F.16 shows mean and 95% confidence interval of the number of repetitions per category for the non-huddle and huddle group. It excludes all suggestions submitted prior to the first huddle, i.e. 14th of January 2020. The categories are defined as follows: 1 Supervisor skills, 2 Leave, 3 Attendance bonus, 4 Work environment (behavioral), 5 Work environment (facilities), 6 Manpower, 7 Machine, 8 Salary, 9 Motion/waiting time, 10 Efficiency, 11 Health and safety, 12 Production layout.

1.G Supplementary tables: Main outcomes

TABLE 1.G.11. LATE (IV) EFFECTS ON PRIMARY OUTCOMES (FIRM PERFORMANCE AND WORKER WELL-BEING)

	Firm performance				**Worker well-being**					
	Workforce index			**Production index**	**Job satisfaction index**			**Mental health index**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Lunch	-0.007		0.035		0.197**		0.933	0.022		0.650
	(0.720)		(0.679)		(0.031)		(0.242)	(0.852)		(0.371)
Huddle		0.050	0.099	-0.332		0.564	1.220		0.846	1.465
		(0.447)	(0.274)	(0.175)		(0.255)	(0.275)		(0.373)	(0.300)
LunchXHuddle			-0.096				-1.232			-1.080
			(0.525)				(0.306)			(0.365)
Observations	119588	92078	91580	459	195	195	195	195	195	195
R^2	0.05	0.05	0.05	0.46	0.20	0.02	-0.13	0.10	-0.10	-0.21
Control Mean	-0.18	-0.13	-0.11	0.15	-0.23	-0.14	-0.15	-0.13	-0.21	-0.12
Clusters	792	55	55	51	182	51	51	182	51	51
Kleib.-P. LM	0.00	0.00	0.00	0.00		0.01	0.09		0.01	0.10
Kleib.-P. Wald F	26920	75	18	219		8	1		6	1
p (L=H)			0.40				0.62			0.43
p (L=LxH)			0.58				0.28			0.37
p (H=LxH)			0.38				0.28			0.30

Notes: Standard errors are clustered at the pair level in columns 1, 5, and 8, and at the line level otherwise. Time fixed effects are included in regressions on the workforce index (in days) and on the production index (in weeks). Huddle exposure is truncated at one (at least one huddle). The cumulative exposure per worker is used in columns 1, 2 and 3, the average cumulative exposure per line in column 4, and the maximum exposure per worker in columns 5 to 10. Indices are constructed using the Anderson (2008) method. Standard errors are in parentheses. * p<0.1, ** p<0.05, *** p<0.01

TABLE 1.G.12. LATE (IV) EFFECTS ON PRIMARY OUTCOMES (INNOVATION)

	Quantity			**Quality**			**Quality (if submitted)**			**Variety**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Lunch	0.61***		0.53***	2.61***		2.21***	-1.17**		-1.25	
	(0.018)		(0.083)	(0.079)		(0.41)	(0.46)		(1.24)	
Huddle		-0.076	-0.078*		-0.32	-0.44		0.057	0.34	0.35
		(0.079)	(0.043)		(0.43)	(0.28)		(0.76)	(2.12)	(0.28)
LunchXHuddle			-0.044			0.051			-0.14	
			(0.12)			(0.60)			(2.09)	
Observations	6352	4956	4956	6710	5179	5179	1857	1294	1294	543
R^2	0.32	0.06	0.27	0.31	0.07	0.27	0.06	0.07	0.07	-0.09
Control Mean	0.03	0.40	0.25	0.08	1.41	0.87	6.02	5.05	5.09	1.18
Round FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Floor FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Kleib.-P. LM	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Kleib.-P. Wald F	8398	66	14	10235	71	16	6239161	33	10	21
p (L=H)			0.00			0.00			0.20	
p (L=LxH)			0.00			0.03			0.73	
p (H=LxH)			0.80			0.50			0.91	

Notes: In columns 1, 4 and 7 standard errors are clustered at the pair level. In all other columns, standard errors are clustered at the line level, and observations prior to the first huddle (14th of January 2020) are excluded. Columns 7 to 10 only include submitted suggestions. Column 10 further restricts observations to individuals who submitted more than once (which are only some of the lunch workers). The quality index (column 4 to 9) consists of six components, out of which two are criteria to evaluate the usefulness of submitted ideas for workers, namely impact on worker well-being and implementation level, and four are criteria to evaluate the usefulness of ideas for the firm, namely financial resources, impact on efficiency, implementation time and risk, and level of effects (separate regressions shown in Table 1.G.14). * $p<0.1$, ** $p<0.05$, *** $p<0.01$

TABLE 1.G.13. EFFECTS ON SUB-COMPONENTS (FIRM PERFORMANCE AND WORKER WELL-BEING)

Panel A: ITT Effects of lunch treatment

	Workforce index				Production index			Job satisfaction index			Mental health index		
	(1) Absenteeism	(2) Punctuality	(3) Working time	(4) Exits*	(5) Efficiency	(6) Alterations	(7) Rejects	(8) Job satisfaction	(9) Quit intentions	(10) Job attitudes	(11) Feelings at work	(12) General anxiety	(13) Harrassment
Lunch	0.001	0.012	0.005	0.001				0.129	0.035	0.362**	-0.048	0.060	0.212*
	(0.002)	(0.224)	(0.025)	(0.015)				(0.128)	(0.083)	(0.151)	(0.104)	(0.141)	(0.121)
	{1.000}	{1.000}	{1.000}	{1.000}				{0.459}	{0.811}	{0.055}	{0.814}	{0.814}	{0.329}
Observations	119588	91591	91747	1450	447	459	459	195	195	195	195	195	194
Clusters	792	788	788	784	51	51	51	182	182	182	182	182	181
R^2	0.02	0.11	0.21	0.03	0.46	0.46	0.24	0.14	0.06	0.27	0.01	0.27	0.06
Control mean	0.03	-3.96	-11.80	0.09	0.59	-0.08	-0.00	-0.44	0.11	-0.16	-0.09	-0.09	-0.09
Time FE	YES	YES	YES	NO	YES	YES	YES	NO	NO	NO	NO	NO	NO
Floor FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES

Panel B: ITT Effects of lunch and huddle treatments

	Workforce index				Production index			Job satisfaction index			Mental health index		
	(1) Absenteeism	(2) Punctuality	(3) Working time	(4) Exits*	(5) Efficiency	(6) Alterations	(7) Rejects	(8) Job satisfaction	(9) Quit intentions	(10) Job attitudes	(11) Feelings at work	(12) General anxiety	(13) Harrassment
Lunch	0.003	0.028	0.000	0.000				0.142	0.136	0.462**	-0.052	0.215	0.112
	(0.002)	(0.263)	(0.020)	(0.025)				(0.164)	(0.164)	(0.183)	(0.122)	(0.201)	(0.163)
	{1.000}	{1.000}	{1.000}	{1.000}				{0.378}	{0.378}	{0.047}	{1.000}	{1.000}	{1.000}
Huddle	0.003	-0.291	0.074*	0.002	-0.005	-0.007*	-0.000	0.393*	0.172	0.031	0.167	0.208	-0.081
	(0.003)	(0.284)	(0.038)	(0.023)	(0.014)	(0.004)	(0.000)	(0.216)	(0.150)	(0.201)	(0.161)	(0.206)	(0.196)
	{0.576}	{0.576}	{0.297}	{0.950}	{0.909}	{0.363}	{0.909}	{0.296}	{0.351}	{0.638}	{0.916}	{0.916}	{0.916}
LunchXHuddle	-0.005	-0.027	0.007	0.002				-0.087	-0.225	-0.203	-0.018	-0.336	0.208
	(0.004)	(0.359)	(0.042)	(0.031)				(0.223)	(0.185)	(0.307)	(0.246)	(0.253)	(0.256)
	{1.000}	{1.000}	{1.000}	{1.000}				{1.000}	{1.000}	{1.000}	{1.000}	{1.000}	{1.000}
Observations	119588	91591	91747	1450	447	459	459	195	195	195	195	195	194
Clusters	51	51	51	51	51	51	51	49	49	49	49	49	49
R^2	0.02	0.11	0.21	0.03	0.46	0.46	0.24	0.17	0.07	0.27	0.02	0.27	0.06
Control mean	0.03	-3.82	-11.90	0.09	0.59	-0.08	-0.00	-0.61	0.03	-0.21	-0.17	-0.19	-0.07
Time FE	YES	YES	YES	NO	YES	YES	YES	NO	NO	NO	NO	NO	NO
Floor FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
p (L=H)	0.83	0.31	0.10	0.92				0.22	0.75	0.08	0.18	0.98	0.36
q (L=H)	1.00	0.88	0.66	1.00				0.31	0.50	0.31	1.00	1.00	1.00
p (L=LxH)	0.18	0.92	0.89	0.98				0.53	0.29	0.14	0.92	0.20	0.80
q (L=LxH)	1.00	1.00	1.00	1.00				0.71	0.71	0.71	1.00	1.00	1.00
p (H=LxH)	0.26	0.65	0.36	0.99				0.22	0.22	0.63	0.61	0.18	0.50
q (H=LxH)	1.00	1.00	1.00	1.00				0.49	0.49	0.49	1.00	1.00	1.00

Notes: Panel A shows ITT effects of the lunch treatment, and panel B of the lunch and the huddle treatments. Effects on the production index can solely be displayed for the huddle treatment due to unavailability of production data at the worker-level. Variables *Exits* (column 4), denoted with a star, is collapsed at the individual worker-level and not included in the workforce index. Standard errors are clustered at the pair level in panel A and at the line level in panel B. False discovery rate (FDR) adjusted p-values, also known as q-values, were used to correct for multiple hypothesis testing. They are shown in curly brackets. These were calculated following the method of Benjamini et al. (2006). Standard errors are shown in parentheses. * p<0.1, ** p<0.05, *** p<0.01

TABLE 1.G.14. EFFECTS ON QUALITY INDEX COMPONENTS

	Worker comp.			**Firm comp.**		
	(1)	(2)	(3)	(4)	(5)	(6)
Lunch	-0.61		-0.74	-1.31***		-1.40*
	(0.42)		(0.63)	(0.50)		(0.74)
Huddle		-0.063	-0.13		0.036	0.28
		(0.19)	(0.89)		(0.19)	(1.18)
LunchXHuddle			0.092			-0.20
			(0.87)			(1.16)
Observations	1857	1294	1294	1857	1294	1294
R^2	0.06	0.08	0.08	0.06	0.06	0.07
Control Mean	6.00	5.44	5.45	6.46	5.24	5.27
Round FE	YES	YES	YES	YES	YES	YES
Floor FE	YES	YES	YES	YES	YES	YES
p (L=H)			0.35			0.07
p (L=LxH)			0.55			0.49
p (H=LxH)			0.90			0.84

Notes: Table 1.G.14 shows individual regressions for worker and firm components fo the quality index. Worker components include the two criteria impact on worker well-being and implementation level. Firm components include the four criteria financial resources, impact on efficiency, implementation time and risk, and level of effects (direct or indirect). * p<0.1, ** p<0.05, *** p<0.01

TABLE 1.G.15. EFFECTS ON DIFFERENT SPECIFICATIONS OF VARIETY

	Categorization			**Latent Semantic Analysis**
	(1) Category	(2) Subcategory	(3) Suggestion	(4) Suggestion text
Huddle	0.022	0.063	0.070	0.010
	(0.058)	(0.059)	(0.052)	(0.014)
Observations	543	543	543	543
R^2	0.03	0.05	0.05	0.02
Control Mean	0.88	0.94	1.11	0.89
Round FE	YES	YES	YES	YES
Floor FE	YES	YES	YES	YES

Notes: Variety in columns 1 to 3 is composed of the number of new categories, subcategories or suggestions resp. of each suggestion paper in comparison with the ones mentioned in any prior paper by the same worker. Variety in column 4 is composed of the percentage of new content based on a latent semantic analysis approach which compares word vectors of each suggestion paper with the ones of any prior paper by the same worker.

TABLE 1.G.16. EFFECTS ON THE CONTENT OF SUGGESTIONS

Panel (a): ITT Effects of lunch treatment

	(1) Supervisor skills	(2) Leave	(3) Attend. bonus	(4) WE (behav.)	(5) WE (facil.)	(6) Manpower	(7) Machine	(8) Salary	(9) Motion waiting	(10) Effi-ciency	(11) Health & safety	(12) Prod. layout
Lunch	-0.056	-0.18***	0.074	-0.095	0.10***	0.046	0.0036	0.035***	0.026***	0.020***	0.022***	0.00035
	(0.060)	(0.070)	(0.049)	(0.063)	(0.040)	(0.032)	(0.028)	(0.0083)	(0.0073)	(0.0073)	(0.0067)	(0.00058)
N	1015	1015	1015	1015	1015	1015	1015	1015	1015	1015	1015	1015
R^2	0.05	0.06	0.02	0.02	0.06	0.05	0.05	0.01	0.02	0.02	0.01	0.01
Ctrl Mean	0.19	0.32	0.11	0.27	0.03	0.03	0.05	0.00	0.00	0.00	0.00	0.00
Round FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Floor FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES

Panel (b): ITT Effects of huddle treatment

	(1) Supervisor skills	(2) Leave	(3) Attend. bonus	(4) WE (behav.)	(5) WE (facil.)	(6) Manpower	(7) Machine	(8) Salary	(9) Motion waiting	(10) Effi-ciency	(11) Health & safety	(12) Prod. layout
Huddle	-0.042*	-0.0077	0.012	0.0063	0.031	0.015	0.014	-0.015	-0.013**	-0.00075	0.0031	-0.0019
	(0.022)	(0.023)	(0.022)	(0.023)	(0.021)	(0.020)	(0.012)	(0.010)	(0.0063)	(0.0087)	(0.0081)	(0.0020)
N	1015	1015	1015	1015	1015	1015	1015	1015	1015	1015	1015	1015
R^2	0.05	0.05	0.02	0.02	0.06	0.05	0.05	0.01	0.02	0.02	0.00	0.01
Ctrl Mean	0.20	0.16	0.17	0.15	0.09	0.08	0.05	0.04	0.02	0.02	0.02	0.00
Round FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Floor FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES

Panel (c): ITT Effects of lunch and huddle treatment

	(1) Supervisor skills	(2) Leave	(3) Attend. bonus	(4) WE (behav.)	(5) WE (facil.)	(6) Manpower	(7) Machine	(8) Salary	(9) Motion waiting	(10) Effi-ciency	(11) Health & safety	(12) Prod. layout
Lunch	-0.035	-0.15**	0.060	-0.11	0.085	0.0043	0.020	0.046***	0.031***	0.022**	0.020**	0.0014
	(0.075)	(0.072)	(0.055)	(0.077)	(0.055)	(0.047)	(0.022)	(0.011)	(0.011)	(0.0095)	(0.0096)	(0.0016)
Huddle	0.0026	0.088	-0.026	-0.020	-0.015	-0.10*	0.066	0.010	-0.0037	0.0041	-0.0021	0.00028
	(0.12)	(0.14)	(0.088)	(0.13)	(0.059)	(0.055)	(0.067)	(0.0092)	(0.0098)	(0.012)	(0.0099)	(0.0011)
LunchXHuddle	-0.045	-0.095	0.038	0.030	0.045	0.12*	-0.054	-0.027	-0.0098	-0.0054	0.0049	-0.0022
	(0.12)	(0.14)	(0.093)	(0.13)	(0.054)	(0.067)	(0.068)	(0.016)	(0.012)	(0.017)	(0.013)	(0.0025)
N	1015	1015	1015	1015	1015	1015	1015	1015	1015	1015	1015	1015
R^2	0.05	0.06	0.02	0.02	0.07	0.06	0.05	0.01	0.02	0.02	0.01	0.01
Ctrl Mean	0.20	0.17	0.16	0.15	0.09	0.07	0.05	0.04	0.02	0.02	0.02	0.00
Round FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Floor FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
p (L=H)	0.68	0.06	0.18	0.44	0.00	0.00	0.49	0.00	0.00	0.06	0.02	0.40
p (L=LxH)	0.96	0.78	0.88	0.46	0.70	0.29	0.34	0.01	0.07	0.26	0.47	0.37
p (H=LxH)	0.84	0.53	0.72	0.85	0.59	0.07	0.38	0.13	0.77	0.73	0.74	0.44

Notes: Panel (a) shows lunch treatment effects, panel (b) huddle treatment effects, and panel (c) lunch and huddle treatment effects on the content of suggestions. Regressions in panel (b) and (c) exclude all suggestions prior to the first day a huddle took place (i.e. 14th of January 2020). Categories are only defined for type four and five (problems and ideas).

1.H Supplementary tables: Mechanisms

TABLE 1.H.17. LEARNING OVER TIME

	Quantity		Quality		Variety
	(1)	(2)	(3)	(4)	(5)
Round2	0.030***	0.030***	0.20***	0.20***	
	(0.0093)	(0.0092)	(0.061)	(0.061)	
Round3	0.0048	0.0050	0.033	0.034	-0.11*
	(0.0052)	(0.0052)	(0.033)	(0.033)	(0.057)
Round4		0.000033		0.018	0.081
		(0.0061)		(0.045)	(0.072)
LunchXRound1	0.70***	0.71***	3.23***	3.73***	
	(0.020)	(0.034)	(0.13)	(0.24)	
LunchXRound2	0.51***	0.61***	2.35***	3.18***	
	(0.023)	(0.038)	(0.15)	(0.26)	
LunchXRound3	0.53***	0.51***	2.51***	2.26***	
	(0.021)	(0.040)	(0.13)	(0.23)	
LunchXRound4		0.57***		3.25***	
		(0.037)		(0.25)	
Observations	4401	4104	4401	4104	542
R^2	0.44	0.51	0.31	0.40	0.04
Control Mean	4.40	0.03	4.40	0.03	1.44
Floor FE	YES	YES	YES	YES	YES
p (LunchXRound1=LunchXRound2)	0.00	0.02	0.00	0.09	
p (LunchXRound2=LunchXRound3)	0.44	0.04	0.36	0.00	
p (LunchXRound3=LunchXRound4)		0.29		0.00	
p (Round3=Round4)					0.01

Notes: Quantity is the number of suggestions per worker per round. Quality is the quality index on a scale from one to ten with non-submissions coded as zero. Variety is the number of categories per worker per round, which have not been mentioned by that worker in any previous rounds. A round is defined as the time between a lunch of sub-group one and the day before the subsequent lunch of sub-group one. Per round, each of the four sub-groups receive one lunch. Columns 1 and 3 include observations of treated workers who were assigned to receive three lunches. Columns 2 and 4 only include observations of treated workers who were assigned to receive four lunches. Column 5 excludes observations from workers who did not submit more than one suggestion, and hence, only includes the lunch group. It further excludes observations from the first round, as variety can only be measured if at least one suggestion per worker has already been submitted. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

TABLE 1.H.18. EFFECTS ON SECONDARY OUTCOME COMPONENTS

	Operators								Supervisors	
	(1) Control μ	(2) Lunch	(3) Control μ	(4) Huddle	(5) Control μ	(6) Lunch	(7) Huddle	(8) Both	(9) Control μ	(10) Huddle
Index: Engagement and support	-0.07	0.12	-0.01	0.03	-0.07	0.07	-0.03	0.09	-0.06	-0.10
		(0.11)		(0.11)		(0.16)	(0.20)	(0.29)		(0.18)
Idea generation	-0.09	0.14	-0.07	0.13	-0.13	0.10	0.08	0.07	-0.12	0.21
		(0.12)		(0.12)		(0.16)	(0.22)	(0.30)		(0.20)
Leadership efficacy	-0.03	-0.05	0.03	-0.01	-0.03	0.09	0.16	-0.31	0.02	-0.28
		(0.15)		(0.12)		(0.17)	(0.17)	(0.23)		(0.18)
Submission of suggestions	-0.06	0.13	-0.01	0.04	-0.04	0.08	-0.02	0.10	-0.05	0.06
		(0.15)		(0.14)		(0.17)	(0.19)	(0.26)		(0.18)
Submission of suggestions (all)	-0.03	0.41***	0.01	-0.07	0.01	0.47***	-0.01	-0.13		
		(0.06)		(0.07)		(0.08)	(0.07)	(0.10)		
Index: Perceived constraints	-0.01	-0.05	0.06	0.11	0.09	-0.18	-0.03	0.28	0.03	-0.03
		(0.11)		(0.12)		(0.17)	(0.20)	(0.25)		(0.29)
Constraints operators	-0.00	-0.10	0.04	0.16	0.06	-0.18	0.09	0.16	0.01	0.07
		(0.13)		(0.10)		(0.20)	(0.17)	(0.28)		(0.26)
Constraints supervisors									0.03	-0.10
										(0.28)
Ease of idea generation index	-0.04	0.01	-0.04	0.06	-0.04	0.03	0.10	-0.07		
		(0.13)		(0.12)		(0.17)	(0.16)	(0.26)		
Ease of submission	0.01	-0.02	0.06	-0.16	0.13	-0.13	-0.30	0.27		
		(0.14)		(0.15)		(0.15)	(0.20)	(0.26)		
Ease of submission (all)	0.30	0.19***	0.34	-0.05	0.40	0.14**	-0.12	0.11		
		(0.06)		(0.06)		(0.07)	(0.09)	(0.11)		
Index: Aspirations	-0.04	0.34***	-0.01	0.09	-0.03	0.29	0.01	0.11	-0.13	0.36*
		(0.12)		(0.12)		(0.18)	(0.16)	(0.18)		(0.20)
Belief about winning and implementing	-0.04	0.34***	-0.01	0.09	-0.03	0.29	0.01	0.11	-0.08	0.28
		(0.12)		(0.12)		(0.18)	(0.16)	(0.18)		(0.20)
Assessment of workers									-0.08	0.26
										(0.16)
General perceptions									-0.15	0.30
										(0.20)
Index: Communication intensity	-0.02	-0.02	0.04	0.00	-0.02	-0.02	0.00	0.00	-0.05	-0.01
		(0.05)		(0.05)		(0.07)	(0.07)	(0.09)		(0.10)
General communication	-0.01	0.06	0.04	-0.02	0.07	-0.00	-0.09	0.13	0.04	-0.28
		(0.15)		(0.15)		(0.17)	(0.17)	(0.30)		(0.17)
Raising suggestions	-0.10	-0.00	0.02	0.01	-0.11	-0.03	-0.02	0.05		
		(0.07)		(0.08)		(0.11)	(0.07)	(0.16)		
Suggestion box communication	-0.06	0.14	-0.03	0.02	-0.13	0.27	0.15	-0.28	-0.20	0.42**
		(0.13)		(0.10)		(0.20)	(0.19)	(0.32)		(0.18)
Suggestion box communication (all)	0.05	0.18***	0.07	0.14*	-0.02	0.21***	0.15	-0.06		
		(0.06)		(0.07)		(0.07)	(0.10)	(0.13)		
Communication with managers									-0.06	0.33**
										(0.16)
Index: Workplace culture	0.04	-0.03	0.06	-0.18	0.11	-0.17	-0.34	0.33	0.01	0.02
		(0.14)		(0.13)		(0.17)	(0.21)	(0.30)		(0.14)
Voice climate workers	0.01	0.05	0.08	-0.23*	0.10	0.01	-0.30*	0.13	0.01	-0.03
		(0.13)		(0.13)		(0.17)	(0.17)	(0.27)		(0.18)
Voice climate supervisors									-0.06	0.09
										(0.23)
Caring culture	0.04	-0.07	-0.04	0.05	-0.05	-0.11	0.01	0.09		
		(0.11)		(0.11)		(0.13)	(0.17)	(0.21)		
Mindset about authority	0.04	-0.13	0.05	-0.07	0.11	-0.29*	-0.24	0.35	0.07	-0.07
		(0.14)		(0.19)		(0.15)	(0.24)	(0.24)		(0.21)
Number of Observations	200		200		200				97	
Number of Observations (all)	1164		1171		1164					

Notes: Standard errors are clustered at the pair level for regressions in columns 1 and 2, and at the line level for all other regressions. * p<0.1, ** p<0.05, *** p<0.01

TABLE 1.H.19. LATE (IV) EFFECTS ON SECONDARY OUTCOME COMPONENTS

	Operators							Supervisors		
	(1) Control μ	(2) Lunch	(3) Control μ	(4) Huddle	(5) Control μ	(6) Lunch	(7) Huddle	(8) Both	(9) Control μ	(10) Huddle
Index: Engagement and support	-0.07	0.12	-0.10	0.14	-0.15	-0.15	-0.16	0.44	-0.06	-0.25
		(0.11)		(0.62)		(0.81)	(1.08)	(1.36)		(0.44)
Idea generation	-0.10	0.14	-0.01	0.67	-0.07	-0.02	0.50	0.22	-0.12	0.54
		(0.12)		(0.74)		(0.90)	(1.26)	(1.53)		(0.48)
Leadership efficacy	-0.03	-0.05	0.01	-0.03	-0.04	0.83	0.83	-1.44	0.02	-0.74
		(0.15)		(0.68)		(0.59)	(1.04)	(1.00)		(0.45)
Submission of suggestions	-0.06	0.13	-0.06	0.23	-0.24	-0.15	-0.08	0.45	-0.05	0.14
		(0.14)		(0.81)		(0.77)	(1.04)	(1.26)		(0.44)
Submission of suggestions (all)	-0.03	0.41***	-0.05	-0.32	-0.22	0.86***	-0.04	-0.66		
		(0.06)		(0.29)		(0.33)	(0.34)	(0.50)		
Index: Perceived constraints	-0.05	-0.05	-0.09	0.49	-0.23	-0.74	-0.03	1.08	0.03	-0.09
		(0.11)		(0.48)		(0.63)	(0.78)	(0.96)		(0.73)
Constraints operators	-0.04	-0.10	-0.14	0.71	-0.26	-0.54	0.41	0.68	0.01	0.18
		(0.12)		(0.54)		(0.72)	(0.63)	(1.04)		(0.65)
Constraints supervisors									0.03	-0.25
										(0.72)
Ease of idea generation	-0.03	0.01	0.07	0.32	0.09	0.19	0.51	-0.32		
		(0.13)		(0.70)		(0.76)	(0.90)	(1.22)		
Ease of submission	0.01	-0.02	0.13	-0.86	0.15	-0.86	-1.69	1.42		
		(0.14)		(0.80)		(0.89)	(1.27)	(1.44)		
Ease of submission (all)	0.30	0.19***	0.40	-0.24	0.42	-0.12	-0.56	0.51		
		(0.06)		(0.28)		(0.31)	(0.43)	(0.50)		
Index: Aspirations	-0.04	0.34***	-0.11	0.48	-0.10	0.05	0.09	0.45	-0.13	0.95*
		(0.12)		(0.65)		(0.52)	(0.89)	(0.81)		(0.51)
Belief about winning and implementing	-0.04	0.34***	-0.11	0.48	-0.10	0.05	0.09	0.45	-0.08	0.74
		(0.12)		(0.65)		(0.52)	(0.89)	(0.81)		(0.50)
Assessment of workers									-0.08	0.67*
										(0.41)
General perceptions									-0.15	0.76
										(0.48)
Index: Communication intensity	-0.03	-0.02	-0.02	0.01	-0.10	-0.03	0.01	0.01	-0.05	-0.03
		(0.05)		(0.24)		(0.27)	(0.36)	(0.43)		(0.25)
General communication	-0.02	0.06	-0.10	-0.09	-0.06	-0.32	-0.49	0.63	0.04	-0.71*
		(0.15)		(0.79)		(0.78)	(1.01)	(1.27)		(0.42)
Raising suggestions	-0.11	-0.00	0.03	0.04	-0.15	-0.14	-0.09	0.23		
		(0.07)		(0.36)		(0.42)	(0.40)	(0.68)		
Suggestion box communication	-0.06	0.14	0.01	0.09	0.01	0.95	0.81	-1.33	-0.20	1.07**
		(0.13)		(0.68)		(0.90)	(1.10)	(1.46)		(0.44)
Suggestion box communication (all)	0.05	0.19***	0.12	0.68**	0.11	0.34	0.75	-0.30		
		(0.06)		(0.29)		(0.38)	(0.52)	(0.61)		
Communication with managers									-0.06	0.87**
										(0.42)
Index: Workplace culture	0.03	-0.03	0.05	-0.97	0.12	-1.05	-1.98	1.73	0.01	0.05
		(0.14)		(0.86)		(0.94)	(1.47)	(1.58)		(0.35)
Voice climate workers	0.01	0.05	0.12	-1.29	0.19	-0.37	-1.77	0.77	0.01	-0.07
		(0.13)		(0.87)		(0.89)	(1.31)	(1.44)		(0.44)
Voice climate supervisors									-0.06	0.24
										(0.58)
Caring culture	0.04	-0.07	-0.15	0.25	-0.17	-0.33	0.05	0.41		
		(0.10)		(0.54)		(0.60)	(0.94)	(0.99)		
Mindset about authority	0.03	-0.13	0.00	-0.39	0.04	-1.16	-1.30	1.70	0.07	-0.19
		(0.14)		(0.74)		(0.80)	(1.42)	(1.39)		(0.51)
Number of Observations	200		200		200				97	
Number of Observations (all)	1164				1164					

Notes: Standard errors are clustered at the pair level for regressions in columns 1 and 2, and at the line level for all other regressions. * p<0.1, ** p<0.05, *** p<0.01

TABLE 1.H.20. HETEROGENEOUS TREATMENT EFFECTS BY WORKERS' SKILLS

	Quantity			**Quality**		
	(1)	(2)	(3)	(4)	(5)	(6)
High-skilled worker	0.026	0.13***	0.014	0.11	0.39***	0.035
	(0.017)	(0.036)	(0.022)	(0.067)	(0.14)	(0.10)
Lunch	0.59***		0.44***	2.29***		1.87***
	(0.026)		(0.036)	(0.10)		(0.19)
Lunch * High-skilled worker	0.068*		0.16***	0.16		0.43*
	(0.038)		(0.056)	(0.15)		(0.25)
Huddle		0.020	-0.012		0.016	-0.064
		(0.030)	(0.015)		(0.12)	(0.051)
Huddle * High-skilled worker		-0.081*	-0.012		-0.15	0.010
		(0.041)	(0.022)		(0.18)	(0.11)
Lunch * Huddle			0.054			0.12
			(0.052)			(0.22)
Lunch * Huddle * High-skilled worker			-0.13*			-0.27
			(0.067)			(0.30)
Observations	4680	3634	3634	4680	3634	3634
R^2	0.31	0.07	0.27	0.28	0.06	0.24
Control Mean	0.01	0.25	0.02	0.04	1.02	0.08
Round FE	YES	YES	YES	YES	YES	YES
Floor FE	YES	YES	YES	YES	YES	YES
Lunch=Lunch*High-skilled worker	0.00		0.00	0.00		0.00
Huddle=Huddle*High-skilled worker		0.13	0.98		0.55	0.58

Notes: Table 1.H.20 corresponds to Table 1.8, but uses the number of processes a worker can perform instead of the experience index as a measure of a worker's skillset. "High-skilled worker" means the respective worker can perform an above-median number of processes, as measured in the skills matrices data. This data is not available for 26% of the workers, explaining the lower number of observations compared to Table 1.8. Columns 2, 3, 5, and 6 exclude all observations prior to the start date of huddles, i.e. the 14th of January 2020. A the bottom, I show p-values from comparing each treatment and treatment interaction against each other. Regressions include round and floor fixed effects. Standard errors are clustered at the pair level in columns 1 and 4, and at the line level in columns 2, 3, 5 and 6. Standard errors are shown in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

TABLE 1.H.21. HETEROGENEOUS TREATMENT EFFECTS BY WORKERS' OWN GENDER AND EXPERIENCE

	Quantity			**Quality**		
	(1)	(2)	(3)	(4)	(5)	(6)
Female worker	0.0095	0.016	-0.035	0.074**	0.072	-0.23*
	(0.0091)	(0.011)	(0.030)	(0.035)	(0.046)	(0.13)
High-exp worker		0.0035			0.00050	
		(0.012)			(0.051)	
High-exp worker * Female		-0.0097			0.015	
		(0.014)			(0.059)	
Lunch	0.61***	0.53***		2.46***	2.10***	
	(0.033)	(0.057)		(0.14)	(0.24)	
Lunch * Female worker	-0.071*	-0.076		-0.47***	-0.43*	
	(0.036)	(0.060)		(0.15)	(0.25)	
Lunch * High-exp worker		0.12*			0.56*	
		(0.066)			(0.29)	
Lunch * Female worker * High-exp worker		0.066			0.13	
		(0.075)			(0.31)	
Huddle			-0.00058			-0.032
			(0.043)			(0.18)
Huddle * Female worker			-0.022			-0.025
			(0.046)			(0.19)
Observations	6352	6352	4955	6352	6352	4955
R^2	0.27	0.28	0.06	0.24	0.25	0.05
Control Mean	0.01	0.01	0.30	0.04	0.04	1.27
Round FE	YES	YES	YES	YES	YES	YES
Floor FE	YES	YES	YES	YES	YES	YES
Lunch=Lunch*Female worker	0.00	0.00		0.00	0.00	
Huddle=Huddle*Female worker			0.81			0.99

Notes: Table 1.H.21 shows heterogeneous treatment effects by worker's own gender and experience. "Female worker" is an indicator to one if the respective worker is female. Experience is measured by an Anderson (2008) index of the number of factories worked at, grade, tenure in the factory, tenure in the industry and the educational level. "High-exp worker" means that the respective worker has an above-median experience index. Regressions in columns 3 and 6 exclude observations prior to the start date of huddles, i.e. the 14th of January 2020. All regressions include round and floor fixed effects. Standard errors are clustered at the pair level in columns 1, 2, 4 and 5, and at the line level in columns 3 and 6. They are shown in parentheses. * p<0.1, ** p<0.05, *** p<0.01

TABLE 1.H.22. HETEROGENEOUS TREATMENT EFFECTS BY EXPERIENCE AND FRIENDSHIP

	Quantity		**Quality**	
	(1)	(2)	(3)	(4)
Friend	-0.0049	-0.093	0.21	-0.14
	(0.047)	(0.083)	(0.20)	(0.36)
High-exp worker		0.28***		0.96***
		(0.060)		(0.23)
High-exp partner		0.12*		0.39
		(0.061)		(0.24)
High-exp worker * Friend		0.010		-0.15
		(0.13)		(0.55)
High-exp partner * Friend		0.31**		0.88*
		(0.12)		(0.50)
High-exp worker * High-exp partner		-0.17**		-0.50
		(0.083)		(0.31)
High-exp worker * High-exp partner * Friend		-0.20		0.19
		(0.19)		(0.76)
Observations	2102	2102	2102	2102
R^2	0.05	0.08	0.06	0.08
Control Mean	0.83	0.66	3.15	2.55
Round FE	YES	YES	YES	YES
Floor FE	YES	YES	YES	YES
High-exp worker=High-exp partner		0.00		0.01
High-exp worker=High-exp worker*Friend		0.10		0.10
High-exp partner=High-exp partner*Friend		0.23		0.45
High-exp p.*Friend=High-exp w.*High-exp p.*Friend		0.07		0.55

Notes: Table 1.H.22 shows the effect of workers' own and their opposite seating partners' experience and friendship on the quantity and quality of suggestions. Experience is measured by an Anderson (2008) index of the number of factories worked at, grade, tenure in the factory, tenure in the industry and the educational level. "High-exp worker" means that the respective worker has an above-median experience index. "High-exp partner" means that the respective worker is seated opposite a worker with an above-median experience index. "Friend" is an indicator equal to one if the workers is seated opposite a person the worker talks to most about work-related issues (at baseline). At the bottom, I show p-values from comparing the effect of different interactions. Regressions exclude observations from workers not assigned to the lunch treatment. They include round and floor fixed effects. Standard errors are clustered at the pair level and shown in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

TABLE 1.H.23. HETEROGENEOUS TREATMENT EFFECTS BY NEW SEATING PARTNERS

	Quantity		**Quality**	
	(1)	(2)	(3)	(4)
New partner	0.042	0.073	-0.095	0.079
	(0.053)	(0.058)	(0.22)	(0.23)
New partner * Previous friend		-0.26		-1.47
		(0.27)		(1.19)
Observations	2760	2760	2760	2760
R^2	0.15	0.15	0.11	0.12
Control Mean	0.87	0.86	3.09	2.96
Round FE	YES	YES	YES	YES
Floor FE	YES	YES	YES	YES
New partner=New partner*Previous friend		0.25		0.21

Notes: Table 1.H.23 shows heterogeneous treatment effects by workers assigned to a new seating partner, a non-friend, after two lunch sessions. "New partner" is an indicator equal to one if a workers is seated opposite a new seating partner in lunch session three. "Previous friend" is an indicator equal to one if a worker has been seated opposite a friend in lunch sessions one and two. I use a Difference-In-Difference specification to detect differences between rounds 1+2 and rounds 3(+4). The table only displays coefficients of the effects in round 3(+4). At the bottom, I show p-values from comparing the effect of a new partner in round three and the interaction with having been seated opposite a friend in previous rounds. Regressions exclude observations from workers not assigned to the lunch treatment. They include round and floor fixed effects. Standard errors are clustered at the pair level and shown in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

TABLE 1.H.24. HETEROGENEOUS TREATMENT EFFECTS BY GENDER DISPOSITION AND WORKERS' EXPERIENCE

	Quantity			Quality		
	(1)	(2)	(3)	(4)	(5)	(6)
High-perc male peers	-0.015	-0.013	-0.030	-0.018	-0.36	-0.28
	(0.040)	(0.087)	(0.17)	(0.16)	(0.35)	(0.64)
Female worker		-0.068	-0.16		-0.77**	-0.84
		(0.093)	(0.18)		(0.37)	(0.62)
Female worker*High-perc male peers		-0.0086	-0.033		0.40	0.035
		(0.098)	(0.18)		(0.39)	(0.67)
High-exp worker			-0.022			0.24
			(0.19)			(0.71)
High-exp worker * Female			0.21			0.29
			(0.20)			(0.75)
High-exp worker * High-perc male peers			0.030			-0.15
			(0.19)			(0.76)
High-exp worker * High-perc male peers*Female worker			0.043			0.68
			(0.20)			(0.82)
Observations	2140	2140	2140	2140	2140	2140
R^2	0.05	0.05	0.07	0.06	0.06	0.08
Control Mean	0.80	0.90	0.73	2.91	3.82	2.68
Round FE	YES	YES	YES	YES	YES	YES
Floor FE	YES	YES	YES	YES	YES	YES

Notes: Table 1.H.24 shows heterogeneous treatment effects by the gender disposition at lunch sessions and worker's experience. "High-perc male peers" means that workers are seated around an above-median percentage of male peers. The percentage of seating partners is measured as the number of male seating partners to the left, right, opposite, diagonally left and diagonally right divided by the total number of seating partners. "Female worker" is an indicator to one if the respective worker is female. Experience is measured by an Anderson (2008) index of the number of factories worked at, grade, tenure in the factory, tenure in the industry and the educational level. "High-exp worker" means that the respective worker has an above-median experience index. Regressions exclude observations from workers not assigned to the lunch treatment. They include round and floor fixed effects. Standard errors are clustered at the pair level and shown in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

TABLE 1.H.25. HETEROGENEOUS TREATMENT EFFECTS ON JOB SATISFACTION

	Job satisfaction			
	(1)	(2)	(3)	(4)
Lunch	-0.26	0.091		
	(0.28)	(0.14)		
Lunch * Above-median submissions	0.83**			
	(0.38)			
Lunch * Above-median experience		0.21		
		(0.18)		
Huddle			0.18	0.13
			(0.13)	(0.14)
Huddle * Above-median submissions			-0.19	
			(0.19)	
Huddle * Above-median experience				-0.039
				(0.19)
Observations	195	195	195	195
R^2	0.22	0.20	0.21	0.18
Control Mean	-0.23	-0.23	-0.17	-0.17
Round FE	NO	NO	NO	NO
Floor FE	YES	YES	YES	YES
p (L=L*Int)	0.08	0.70		
p (H=H*Int)			0.22	0.58

Notes: Table 1.H.25 shows heterogeneous treatment effects on job satisfaction by the number of submitted suggestions and the experience index for both the lunch and the huddle treatment separately. All regressions include floor fixed effects. Standard errors are in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

CHAPTER 2

Does Putting the Worker First Work? Evaluating a Worker-Centred Management Training Programme

with **Christopher Woodruff, Atonu Rabbani and Rocco Macchiavello**

Abstract Factories in lower-income countries are characterized by low efficiency and poor working conditions. Can low-cost management interventions lead to improvements on both fronts at the same time? We conduct a low-cost management training programme for managers and supervisors in 25 ready-made garment factories in Bangladesh. Training is provided in a combination of group-based modules and individual factory "activations". These group-based modules focus on productivity and human resource management practices intended to improve productivity and provide direct benefits to workers. Using both detailed survey data and administrative records, we find the programme has a significant positive effect on workers. Practices aimed at improving communication on the production floor seem to have been more effective than technical and production-related practices. The programme had the strongest effects on human-resource related outcomes, leading to decreases in exit rates, turnover and absenteeism. However, most of the effects disappear within six months of treatment. We see some evidence of spillovers of practices, particularly to production lines on the same floor as the treatment line, but the spillovers do not appear to explain the fade of the treatment effects.

2.1 Introduction

There are large and persistent differences in performance among firms in developing countries (Hsieh and Klenow, 2009; Syverson, 2004). Bloom et al. (2013) show that variation in management practices explain an important part of these differences. A randomised experiment carried out in 17 Indian textile factories showed that improved management practices led to productivity improvements of 17%. The improvement in management practices in Bloom et al.'s experiment were generated by a programme of intensive individualised consulting with a market cost of $250,000 per firm. There is also evidence that consulting programmes of more modest scale and costs can improve management practices and productivity and increase firm size. Iacovone et al. (2018) show that a group-based consulting approach can have similar effects as individual one-on-one consulting at one-third of the cost in a sample of medium-sized automotive parts manufacturing firms in Columbia. Bruhn et al. (2018) show that less intensive and less costly individual consulting led to significant employment growth among micro, small and medium-sized enterprises in Mexico.

It is less clear how the changes in management practices generated by these consulting interventions affect the conditions of workers in the factories. We analyse a consulting intervention that aims to bring working conditions to the front. "Benefits for Business and Workers" (BBW) programme is a group-based consulting programme designed to improve firm productivity through management practices that also improve working conditions. The programme was initially designed for very large, export-oriented garment factories. We use a randomised control trial to assess the impact of BBW in 25 ready made garment factories in Bangladesh. The aim of BBW is to offer training and practical expert advice to key factory managers on productivity and human resource management practices. The training programme consists of 24 days of classroom-based training conducted in groups of six factories (off-site) and four follow-up "activation" visits plus one roll-out session, conducted on-site at each factory individually. The sessions were delivered

over a period of six months, and had a full cost of around GBP 8000 per factory.[1] We assess the impact of the training on productivity. However, in a context where buyers care about social compliance, it is possible for the intervention to be profitable to factories even if productivity is not improved.

A second characteristic of BBW is that the training is designed with a complementarity of practices in mind: Changing one practice may be effective only if another practice is also changed. The consultants worried that managers would implement practices one at a time, and then reject them one at a time as ineffective. Therefore, they asked each factory to select one line as a "pilot line" on which all of the initial effort would be focused. The pilot line approach also allowed factory management to see the benefit of the changes at a lower cost. After *all* of the changes were implemented on the pilot line, and managers saw that as a bundle the practices were effective, the practices would then be rolled out in the entire factory. The group training sessions involved a "vertical slice" of managers from the line manager on the pilot line up through the highest-level manager in the factory, in each of the four departments - production, quality, human resources and industrial engineering - with responsibility for management of the sewing floor. For example, from the production department, the line chief, the floor-in-charge and the production manager were meant to attend the group training sessions. For the purposes of our experiment, we asked factories to nominate three lines as potential pilot lines, and then we randomly selected one of the three as the pilot line, using the other two lines as control lines.

We use detailed administrative data to examine production and HR outcomes, for example, efficiency and worker turnover. We also conducted surveys at three points in time: a baseline prior to the start of training, a midline during the training window and a follow-up after the end of training. The survey data, in particular, allow us to examine changes at a much more detailed level than in previous consulting interventions - in measurement of practices, the measurement of adoption,

[1]The exchange rate at the start of training in September 2015 was around $1.50 per £, implying a full cost of around $12,000 per firm. Around half of the cost was paid by a subsidy provided by the UK Department for International Development. The other half was normally shared roughly equally by the factory and one of their large international buyers. For our experiment, the project paid 75 percent or all of the fee that were usually paid by the factory.

and the effect of practices on workers. This allows us to understand which practices are more easily adopted, illuminating why changing practices is difficult.

The primary dimension of randomisation is across lines within the factory. As noted, we asked factories to nominate three lines as possible pilot lines, and then selected one of those lines at random. However, we need to address the possibility of the spillover of practices across lines. Indeed, some practices are implemented at a level higher than a single production line. For example, the training encouraged the HR team to assign newly-hired workers a "buddy" - an experienced worker who could answer questions about factory practices and act as a mentor. When workers are hired it is often unclear on which line they will work. So the buddy practice was implemented at the production floor or unit level. With these challenges in mind, we asked managers in factories with more than one production unit to nominate three potential pilot lines from at least two different production units. Fewer of the practices will immediately spillover to lines in other units. While some practices are implemented at levels higher than the production line, many of the practices taught in the group sessions were implemented at the line level. In these cases, spillovers are most likely to be seen on lines located adjacent to the pilot line. We survey workers and managers of one of the adjacent lines. The combination of data from the adjacent line, control lines in the same production unit as the pilot line and control lines in a different unit from the pilot line allow us to characterise spillovers.

The project contributes to two strands of the literature. First, it contributes to the growing literature on the experimental evaluation of training and consulting programmes in firms to improve business practices and management. One part of this literature focuses on training programmes for microenterprises (see McKenzie and Woodruff, 2014, for a review). The other part more related to our project focuses on intensive individualised consulting programmes in small and medium-sized enterprises (Bloom et al., 2013, 2018; Bruhn et al., 2018; Higuchi et al., 2017). Second, the project is related to the recent literature on the effect of business networks and group-based training programmes on firm performance (Boudreau, 2020; Chatterji and Fuqua, 2018; Dalton et al., 2018; Iacovone et al., 2018; Lafortune, 2018; Cai and Wang, 2018).

The rest of the paper is organised as follows: Section 2.2 contains the context in the Bangladeshi garment industry and the sample selection, Section 2.3 the experimental design. Section 2.4 discusses the data used. The main empirical specification and results for worker outcomes are presented in Section 2.5, and for factory outcomes in Section 2.6. Section 2.7 concludes.

2.2 Context in Bangladesh and sample selection

2.2.1 The ready-made garment industry

Most ready-made garment factories in Bangladesh specialise in the production of a particular family of garments. Factories participating in this project produced light knit and woven garments such as t-shirts, jeans, and dress shirts. Production in woven and light knit factories is typically organised in three sections: cutting, sewing and finishing. Many larger light-knit factories are also integrated into knitting of the fabric as well. The BBW programme focuses on the sewing section, which usually employs around 60 percent of the workers in the factory.

Sewing is organised in production lines. The cut fabric enters at the top of the line and the finished product exits at the end of the line. Each worker performs a single process – for example, sewing the side seam of a t-shirt. Depending on the complexity of the garment the line may have between 15 and 70 sewing machine operators and five to 15 helpers, although there is considerable variation across factories in the ratio of helper to operators. A typical production floor has eight to ten production lines, though some may have more or fewer. The median factory in our sample has 1,706 employees in the sewing section and 50 employees per sewing line (see Table 2.2).

2.2.2 Management hierarchies

The management structure in Bangladeshi garment factories is in general very hierarchical within each department (see Figure 2.A.4). The main departments involved in our intervention are Production, Industrial Engineering (in most factories), Quality, and HR. The HR department also includes the Welfare Officer. Most of the participating factories have dedicated Industrial Engineering ("IE") departments that manage most technical aspects of production, including setting targets

and determining the layout of the line. Using IE techniques is certainly thought to be best practice, yet not all factories have IE departments.

Supervisory Structure: The supervisory structure in the production department tends to depend upon a variety of factors, most prominently on the number of machines on each production line. A line producing a more complex garment might have 40 machines and will typically have two line supervisors and one line chief who may be in charge of one to three lines. At the other extreme, lines that produce simple garments like tank tops might have 12 machines and one line supervisor working across multiple lines. Some factories do not have line chiefs, and there are other idiosyncratic variations. The duties of line supervisors are various and include motivating workers, teaching workers new production techniques, identifying quality issues and working with Industrial Engineers to solve problems that may cause bottlenecks on the line. Up the production department chain of command is the 'Floor-In-Charge', who is responsible for production on the entire floor. Managers above are responsible for groups of floors and/or the whole factory. There is similar variation in the structure of other departments. For example, quality inspectors are usually responsible for one line, while quality supervisors may be assigned to one or to several lines. In the HR department, responsibilities cover HR processes across the whole factory. In some IE departments, IE Officers are responsible for one specific line, whereas in others their responsibility covers more than one line.

Decision-making process: The factories in our sample are typically highly centralised, with approval from top-level managers required for most decisions. Illustrative of this, we implemented a closed-end version of the World Management Survey instrument among similar factories from a different project. Among 12 multi-location factories, where headquarters is located outside the factory, all 12 said approval from headquarters was required to spend even USD 1200, less than .01% of the typical factory's output. Within factories, lower and middle managers are responsible to act upon and implement the decisions made at higher levels, and to supervise workers. Hence, for decisions involving multiple departments at the floor level there is an "inverted-U" process, with the process requiring the approval

to be carried upward through the chain of each department. Implementation of the decision agreed by the top managers of the several departments then requires for the final decisions made at the top to be carried back downward through each chain to lower managers in each department. As soon as more than one department is involved in decision-making, the complexity of the coordination process to make decisions sharply increases, resulting in delays or failures to make decisions.

2.3 Experimental design

2.3.1 Description of the intervention

The intervention consisted of a training programme called "Benefits for Business and Workers" (BBW). The aim of this management consulting service was to offer training and practical expert advice to key factory managers on productivity and human resource management practices. It was jointly carried out by Impactt Ltd, a UK-based consultancy, and Rajesh Bheda Consulting, an Indian-based consultancy, and was aimed at enhancing the competitiveness in the garment industry. The programme provided training in both technical interventions intended to directly improve productivity and also in human resource interventions to indirectly increase productivity by lowering absenteeism rates and labour turnover rates through higher motivation and wellbeing of workers. As the name of the programme suggests, practices were designed with the aim of providing benefits simultaneously for both workers and the firm. Implementing an attendance bonus policy serves as a good example to illustrate the purpose of the programme. The programme encouraged factories to provide small bonus payments to workers without any unexcused absences. On the one hand, workers benefit financially from such a policy. On the other hand, the firm benefits in terms of higher planning security as workers are motivated to be less absent in an unexcused manner.[2]

Participating factories were organised into groups of six for the classroom component of the training. Modules related to productivity (PR) and human resource (HR) practices were taught during group-based class-room sessions (off-site) on 24 days over a period of roughly six months. The programme also provided four

[2]Workers remain eligible for the bonus if they arrange for allowed leave in advance.

visits by the consulting team to each participating factory for the purpose of "activation" of what was learned in the training sessions. These on-site sessions focused on implementation of the practices on the pilot line, but did not contain any other individualised consulting. Finally, the consulting team visited the factory at the end of the training programme to begin the process of rolling the practices out to the other lines in the factory, initially working with three additional lines chosen by the factory. Factories could continue rolling out the strategies on further lines themselves after the end of the training programme. Most of the practices taught in the programme were not practiced in participating factories before the start of the intervention.

Training content: The aim of the training was to develop solid and robust systems allowing sustained improvements in productivity and quality and, simultaneously, improvements in workers' job quality. The HR modules covered recruitment design, simplification of policies and their implementation, improvement of access to social benefits, improvement of pay and reduction of working hours, staff training, performance management and incentive setting as well as data management. Training also emphasised communication approaches, such as the importance of listening to workers' voice, grievance procedures and feedback systems, or worker committees. The Production modules covered quality, procurement and several practices related to industrial engineering: line balancing, operation standardisation, production planning, as well as data management and analysis. Examples of the HR and production practices are shown in Table 2.1.

During the group training sessions, managers were assigned "homework" that included the review of existing practices, the practical implementation of new practices on pilot lines, the tracking of new metrics, or the evaluation of certain practices or the development of a roll-out plan. These issues were then discussed by the group at the following session. For more information about the training modules see Section 2.A.2 in the Appendix.

Participating factories: A total of 25 factories participated in the project. The majority of factories (17 factories) came from one large buyer, and seven factories came from one of three other buyers. One factory joined through direct contact

TABLE 2.1. EXAMPLES OF HR AND PRODUCTION PRACTICES

HR practices	
Worker induction system	Showing fire exits, Welfare office, etc. to new hires
Buddy system	Assignment of experience operators as mentors to new hires
One-step leave policy	simplifying the process of asking for allowed leave to reduce the time cost for operators
Exit interviews	Conduct of exit interviews with workers who plan to quit their job to figure out potential reasons
Worker participation committee	Informing workers about the functioning of worker participation committees and allowing them to raise concerns to management
Promotion policy	Clarifying promotion policies and supporting workers in attaining the skills required for promotion
Job description	Information about roles and responsibilities, and basis to measure performance
Attendance bonus	Introducing a bonus payment for attendance
Communication approaches	Facilitation techniques for good quality conversations with workers to understand and respond to their needs
Production practices	
Up-skilling	Providing continuous training for workers to improve their skills
Capacity Study	Sheet to provide information on the operator's potential capacity, i.e. average cycle time of an operation
Skills matrix	Tracking of existing manpower skills and skills gaps
Line balancing	Levelling workload to avoid bottlenecks and excess capacity
Non-productive time	Measurement of non-productive time
Zero Defect Initiative	Training selected operators to inspect the quality of garment in process, allowing defects to be addressed before the garment is assembled further

with the research team rather than through a buyer. Buyers were initially contacted in Bangladesh and the UK. Meetings with supplying factories were then organised in Bangladesh, with 37 factories being invited to supplier engagement meetings between June and September 2015. Initially, 27 factories enroled in the programme, but two of these factories dropped out before the start of the programme due to closure/internal problems.

Participants within factories: The training programme is designed to train jointly managers from different departments. In total, between five and seven managers from each factory attended each session. The specific managers sent varied by factory, but factories most commonly sent floor-level ("Officer") managers. A typical group included the line chief and floor-in-charge from Production, the welfare officer from HR/Welfare, a senior (floor-level) IE officer from Industrial Engineering and a quality officer from the Quality department. Factories were grouped in batches of five or six factories so that each classroom session was intended to have between 30 and 40 managers attending. Workers were not directly involved in the training programme.

Costs: The total cost of the programme was about $12,000 per factory. At the time of the project, a grant from UK Aid covered half the cost of the training. The other half was typically split between the buyer and the factory. The UK-based consultancy offered buyers the option to pay a lump sum that then allowed them to enroll supplier factories for a cost of $750 per factory, to be paid by the factory. The fee paid by the factory was seen as a signal of the factory's commitment to the programme. For this project, the $750 fee was waived for the 17 suppliers of the large buyer. The buyer in turn promised to provide support and encouragement to substitute for the financial contribution of the factories. The remaining eight factories paid the participation fee of $750.

2.3.2 Selection of treatment lines and assignment into batches

We designed the project as a randomised controlled trial to provide training to the managers who were in charge of one chosen treatment line per factory. The normal protocol was for the consultant to ask each factory to choose one line as the pilot

line. Managers from that production line and a vertical slice of the hierarchy above that line were then invited to the relevant classroom sessions. As noted above, the vertical typically included floor-level managers from each department, with both the line chief and floor-level manager from the production department. We expected factories to nominate pilot lines based on factors that made them more susceptible to change and adaptation, including open-mindedness and willingness to learn. Hence, the lines selected to be a pilot line may not be comparable to other lines in the factory.

For the purpose of the research project, we asked factories to nominate three potential pilot lines. Factories were asked to nominate lines on at least two different floors where that was possible. We then randomly chose one of the three nominated lines to be the treatment line, with the other two nominated lines designated as control lines. Although the programme is designed to focus attention exclusively on the pilot line initially, we expected the taught practices to spillover to other lines, either because practices (e.g., worker induction) are implemented at the floor level, or because of communication among line managers. We expected these spillovers to be stronger on the same floor than across floors/units. We therefore differentiate between control lines on the same floor and control lines on different floors in the analysis. We define control floors as floors with no treatment lines but at least one control line, and treatment floors as floors with a treatment line (with or without control lines). In two-thirds of the factories (17 factories), there is a control floor, i.e. at least one control line on a different floor than the treatment line. We classify fourteen of those factories, which only have control lines on a different floor than the treatment line, as type "A", and three factories, which have control lines on both the treatment and the control floor, as type "B". In the remaining eight factories, there is no control floor, i.e. selected control lines are only on the same floor as the treatment line. Those are classified as type "C". We conducted surveys on all of the treatment and control lines. Additionally, we surveyed workers and managers on the (up to) two lines adjacent to the treatment line to measure spillovers. In case the control lines were adjacent to the treatment lines and hence, subject to more direct spillovers, they were dropped as control lines (in ten out of 25 factories).

Assignment into batches: The factories were grouped into five batches of five or six factories that received the training programme jointly (see Appendix Section 2.A.5). The start of training for the five batches was staggered over a period of four months.[3] The initial design called for randomising the participating factories across batches, but for logistical reasons, this proved infeasible. The assignment into batches was affected by two factors. First, the composition of batch one was pre-determined: Two non-sample factories and four factories of the main sample were assigned to batch one on a 'first come, first served' basis. Second, the composition of batches two to five was driven by the location of the factories, as we needed the training to be in the general area of the factories. This reduced the flexibility to assign factories completely randomly. Instead, 11 factories which were roughly in the same location were randomly assigned into batches two and three, and five factories in other locations were placed into batch four. Another five factories which decided to join the programme later were placed into batch five.

2.3.3 Selection of survey participants

Selection of surveyed operators: We surveyed five lines from each factory: The treatment line, the two control lines, and the two lines adjacent to the treatment line. From the treatment and control lines, we selected at random three to survey (two from the front half and one from the back half of the line). Additionally, we surveyed the operator on the line who had joined the factory most recently. On the adjacent lines, we selected at random two operators, one from the back half and one from the front half. As a measure of communication between production workers and their supervisors, we asked the randomly selected operators a series of questions about their life and family. For example, we asked how many children they had, and the age of their children. This is the sort of information that is beneficial for the line supervisor to know, because illness of children is a common reason for unexpected absence of workers. One section of the line supervisor/line chief (LS/LC) survey therefore asked respondents to answer questions about an operator, and we allowed the supervisor to select among the three operators surveyed one he knows "better". For this purpose, we gave the names of two out of three

[3]The timing of the batches could not be staggered further because the UK Aid-sponsored programme ended in June 2016 and all training had to be completed by that time.

surveyed experienced operators on the LS/LC's line and allowed the supervisor to choose one of them.

2.4 Data collection and measurement

2.4.1 Data sources

We study the impact of the training programme by combining both the detailed survey data and the factory administrative data (production and salary data). Survey data is used to measure whether workers experienced any changes and are better off as a result of the training, whereas firm data is used to analyse whether the factory experienced any changes and is better off. We carried out surveys at three different points in time: Just before the start of the training (baseline survey), two to three months after the first training (midline survey) and after the end of the training programme (endline survey). The timing is depicted in figure 2.A.1. Respondents to those survey questions include line operators (LO), line supervisors/line chiefs (LSLC) and higher-level managers (MG). For the operator survey, we randomly selected four operators from five lines per factory. We collected production and salary data on all lines in the units participating in the study (units with either treatment or control lines) from 8 weeks prior to the beginning of training until approximately 6 months after the end of the training.

2.4.2 Outcome measures

Worker outcomes (Survey data): To measure effects on employees, we define four indices that align with different practices taught in the programme: Human relations (HR), Production (PR), worker wellbeing and communication. The HR index measures the implementation of HR-related practices, and similarly for the production index. The wellbeing index tracks perceived wellbeing and the communication index measures changes in patterns of communication between line workers and the supervisors. Each index consists of a group of survey questions about specific practices/approaches which are described in detail in Appendix Section 2.B.3. As not all practices are covered in each survey, indices contain different questions for each group of survey respondents (operators, supervisors/line

chiefs and managers). Indices are constructed as a weighted mean of those standardised questions, with questions are weighted by the inverse of the covariance matrix (Anderson, 2008). All of the practices on which the indices are based are introduced in the first or second training module before the midline survey was carried out (see Figure 2.A.2).

Firm outcomes (Factory data): To measure the effect of the training programme on firm outcomes, we construct two indices, one of which aggregates the HR-related outcomes and the other which aggregates the production-related outcomes. The aggregation improves statistical power to detect effects that go in the same direction (based on Kling et al., 2007). The HR index consists of four HR-related outcome variables:

- Exit rates: The percentage of workers quitting the factory in a given month within two months of joining;
- Turnover rates: The percentage of workers with at least two months of tenure quitting the factory in a given month;
- Absenteeism rates: The percentage of work-days missed in a given month;
- Promotion rates: The percentage of workers promoted in a given month;

The PR index consists of three production-related outcome variables:

- Efficiency: The productivity in a given week, measured using the standard measure in the industry;
- Alteration rates: The percentage of altered pieces in a given week;
- Reject rates: The percentage of rejected pieces in a given week;

The efficiency measure is the ratio of output minutes and available minutes. Output minutes is, in turn, the product of the number of pieces of a garment produced and the standard minute value (SMV) of the garment. The industrial engineering department estimates the SMV by identifying each step required to produce the entire garment, and then estimating the number of seconds a fully efficient worker would take to complete each step. The sum of these times yields the SMV for the full garment. For example, a tank top might have an SMV of 4 minutes and a simple t-shirt of 6 minutes. This implies that 20 fully efficient workers would

produce 300 tank tops or 200 t-shirts each hour. The SMV is an essential determinant in price negotiations with buyers. Available minutes is simply the number of production workers (helpers and operators) multiplied by the number of minutes the line operates. To calculate the daily line-productivity measure, daily output is multiplied by the SMV, and then divided by total labour input on that line and day measured in worker-minutes. The other outcome variables are more transparent, but further definitions of these variables can be found in Appendix Section 2.B.2. The HR index is based on monthly salary data; the production index is based on daily production data, aggregated up to the weekly level. We adjust for the fact that we are testing multiple outcome using the 'sharpened q value' approach based on the Benjamini-Hochberg procedure (Benjamini et al., 2006). We report false discovery rate-adjusted q-values for HR-related and PR-related outcome families.

The availability of firm data up to 6 months after the end of the project allows us to not only measure short-run effects against a set of comparison lines that were not selected by the factories to receive training, but also medium-run effects and spillovers to comparison lines.

2.4.3 Descriptives and balance

The experimental firms all produced garments for the international export market. Table 2.2 reports summary statistics for the sewing sections of these firms (as the experiment focuses on the sewing section only). On average, the sewing section has about 1,874 employees. Sewing processes are organised in production lines of on average 54 workers per line with approximately 8 lines per floor. Almost 90% of all workers on the sewing floor are female. Workers have an average tenure in the factory of two years and eight months. Monthly labour turnover is about 6 percent, but exit rates in the first two months after joining are significantly higher, with 24 percent of workers who join leaving within two months. Absenteeism is modest, with 3 percent of workers absent on a typical day; 2 percent of workers are promoted each month, on average. The complexity of a style is proxied by the "standard minute value" (SMV), as defined above. As noted, the daily line-productivity is daily output multiplied by the SMV, and then divided by total labour input on that line and day measured in worker-minutes. By this measure,

TABLE 2.2. SUMMARY STATISTICS

	Mean	SD	Median	Min.	Max.	Obs.
Salary data						
Employees per firm	1873.95	972.45	1706.00	130.00	3707.00	5010
Employees per line	53.61	26.25	50.00	1.00	152.00	4740
Share of female operators	0.88	0.11	0.89	0.00	1.00	4943
Tenure (in months)	32.30	16.20	28.88	1.00	116.84	4948
Exit rate (within two months)	0.24	0.34	0.00	0.00	1.00	4948
Turnover rate	0.06	0.11	0.03	0.00	1.00	4948
Absenteeism rate	0.03	0.03	0.02	0.00	0.49	4332
Promotion rate	0.02	0.08	0.00	0.00	1.00	4948
Production data						
Standard minute value	21.08	12.50	18.62	3.23	60.00	9669
Daily output	4613.91	5955.72	2796.00	0.00	27460.00	24519
Running days	11.64	15.92	6.00	0.00	93.00	17558
Efficiency	0.41	0.17	0.39	0.09	0.96	9354
Alteration rate	0.05	0.06	0.04	0.00	0.40	20733
Reject rate	0.001	0.004	0.000	0.000	0.026	20733

Notes: Table 2.2 shows the summary statistics of key variables from the monthly salary and weekly production data over a time period of three months before the first training and three months after the last training. Exit rate is the share of workers leaving the factory within two months of joining. The turnover rate is the share of workers leaving the factory with a tenure of more than two months. The absenteeism rate is calculated by dividing the number of absent days by the total number of absent and present days per month. The promotion rate measures the share of workers who moved up grade within a given month. The standard minute value proxies the complexity of a style and is the sum of the time, in seconds, it takes to perform each sewing operation to assemble one piece of the style. To calculate daily efficiency per line, daily output is multiplied by the SMV, and then divided by total labour input on that line and day measured in worker-minutes. Running days are the number of days a style is produced per line. To measure quality, we use alteration and reject rates, which are the number of alterations and rejects, respectively, divided by the total number of garments checked.

the average productivity is 41%. Another measure of productive efficiency is the percentage of garments that require alterations, or re-work. The average alteration rate is around 5 percent. A final measure is the percentage of garments that are rejected due to too many defects. The average reject rate is 0.1%.

Balance: We do not observe significant differences between the treatment, adjacent and control group for any of the firm outcomes apart from turnover (see Table 2.4). We regress the treatment indicator (for treatment or adjacent lines) on all outcomes to test for joint significance. The F-test statistic is 3.04 for control and treatment lines, and 5.50 for control and adjacent lines, suggesting that the randomi-

sation produced highly comparable groups. For worker outcomes, the measures at baseline and midline/endline differ in their composition of survey questions. At baseline, only 40% of the questions for operators and 45% of the questions for supervisors/line chiefs are available.[4] In Table 2.3, we show difference in means tests for these available components, and the corresponding index. We do not observe any significant differences between treatment, adjacent and control group for the PR and Wellbeing index for both operators and supervisors/line chiefs, but we find significant differences for the HR and Communication index for operators. We are somewhat less concerned about these imbalances due the varying composition of the index between baseline and midline/endline.

2.4.4 Compliance with the experimental protocol

We measure compliance by examining whether those managers assigned to training actually attended the training. Attendance at the classroom sessions is an indication of engagement by management in the broader consulting initiative. Full compliance with the protocol is often difficult because of communication issues, shocks faced by the factory (e.g. production pressures, absenteeism, etc.), and other factors. However, we observe a relatively stable attendance rate among key attendants, which is slightly decreasing between the second and third training sessions. Attendance averages approximately 70 percent in the earlier sessions, falling to an average of 65 percent in the later sessions (see Figure 2.B.10). Key attendees, those required to attend the training sessions, were mainly the line-level production manager responsible for the pilot line and mid-level managers responsible for several lines or the floor. Higher-level managers, responsible for production units or the whole factory, were expected to attend only the plenary session kicking off the training programme, but were not required to attend the training. Over the course of the training, we can see a decline in the attendance of middle managers in all departments apart from the HR department, whereas lower managers show a fairly stable attendance (see Figure 2.B.12).

[4]Out of 46 questions for operators at midline/endline, 18 are available at baseline. Out of 58 questions for supervisors/line chiefs at midline/endline, 26 are available at baseline.

TABLE 2.3. BALANCE TABLE: WORKER OUTCOMES

	Control	Treatment	Adjacent	se(C-T)	p(C-T=0)	se(C-A)	p(C-A=0)	N
Line Operators								
HR	-0.00	0.22	-0.19	(0.12)	0.08	(0.06)	0.15	310
Induction	0.04	0.08	-0.17	(0.13)	0.75	(0.07)	0.13	310
Paired with operator	-0.10	0.25	-0.09	(0.13)	0.01	(0.08)	0.95	310
Promotion beliefs	0.03	0.09	-0.12	(0.11)	0.56	(0.07)	0.26	310
PR	-0.01	0.11	-0.03	(0.14)	0.41	(0.06)	0.84	310
Efficiency of instructions	-0.01	0.11	-0.03	(0.14)	0.41	(0.06)	0.84	310
Wellbeing	-0.00	0.02	-0.06	(0.12)	0.86	(0.07)	0.67	310
Conflict with managers	0.12	0.06	-0.20	(0.16)	0.71	(0.08)	0.07	306
Self assessment	-0.12	-0.09	0.18	(0.12)	0.77	(0.07)	0.04	310
Peer assessment	-0.06	0.06	-0.02	(0.13)	0.39	(0.07)	0.80	309
Communication	0.00	0.37	0.32	(0.17)	0.03	(0.07)	0.02	310
Initation of meetings	-0.08	0.07	0.02	(0.15)	0.30	(0.07)	0.48	310
Knowledge about supervisor	-0.23	0.15	0.14	(0.14)	0.01	(0.07)	0.01	310
Line Supervisors / Line Chiefs								
HR	-0.03	-0.18	0.17	(0.17)	0.39	(0.07)	0.17	252
Last leave request	-0.03	-0.18	0.17	(0.17)	0.39	(0.07)	0.17	252
PR	-0.00	-0.08	0.11	(0.18)	0.67	(0.09)	0.54	252
File checks	-0.02	-0.15	0.01	(0.18)	0.49	(0.09)	0.84	252
Flags	-0.18	0.06	-0.08	(0.19)	0.22	(0.07)	0.48	252
Stopwatch	-0.15	0.10	0.02	(0.17)	0.14	(0.08)	0.29	252
Efficiency of processes	0.00	-0.31	0.12	(0.14)	0.03	(0.08)	0.43	252
Reduction of movements	0.02	-0.23	0.09	(0.15)	0.09	(0.08)	0.67	252
Bottleneck solutions	0.09	-0.04	-0.00	(0.18)	0.47	(0.09)	0.60	252
Wellbeing	-0.00	0.06	-0.16	(0.19)	0.75	(0.09)	0.35	252
Collaboration	0.02	0.06	-0.19	(0.17)	0.83	(0.08)	0.19	251
Self assessment	-0.02	0.17	-0.11	(0.18)	0.28	(0.07)	0.53	252
Peer assessment	0.00	-0.08	0.11	(0.18)	0.66	(0.08)	0.52	252
Suggestions	0.00	-0.03	-0.05	(0.17)	0.84	(0.08)	0.72	251
Communication	0.00	-0.06	-0.11	(0.16)	0.71	(0.07)	0.46	252
Initiation of meetings	0.06	-0.03	-0.00	(0.15)	0.54	(0.07)	0.67	252
Communication of targets	-0.08	-0.05	-0.09	(0.17)	0.86	(0.07)	0.96	252
Knowledge about operators	0.10	-0.06	0.00	(0.18)	0.39	(0.08)	0.53	252
Performance meetings	-0.03	0.23	-0.04	(0.22)	0.25	(0.10)	0.96	146
Line target meetings	0.00	0.02	-0.12	(0.24)	0.95	(0.11)	0.58	135
Operator target meetings	-0.18	-0.01	-0.04	(0.21)	0.43	(0.11)	0.53	95

Notes: (C-T) indicates the difference of average values of control lines and the average values of treated lines, while (C-A) is the same calculation comparing control lines with adjacent lines, with *se* and *p* showing the standard deviation and p-value of the difference. We use HR, PR, Wellbeing and Communication indices at baseline. Due to a different composition of indices at baseline and midline/endline, we also show individual index components which are available at baseline.

TABLE 2.4. BALANCE TABLE: FIRM OUTCOMES

	Control	Treatment	Adjacent	Non-survey	se(C-T)	p(C-T=0)	se(C-A)	p(C-A=0)	se(C-Ns)	p(C-Ns=0)	N
Salary Data											
Share of female operators	0.87	0.83	0.83	0.90	(0.03)	0.24	(0.01)	0.24	(0.01)	0.04	1124
Tenure (in months)	36.90	38.70	37.83	33.74	(5.56)	0.75	(2.60)	0.86	(1.40)	0.42	1127
Exit rate	0.20	0.16	0.16	0.21	(0.06)	0.57	(0.03)	0.45	(0.02)	0.86	1127
Turnover rate	0.08	0.04	0.05	0.08	(0.02)	0.08	(0.01)	0.11	(0.01)	0.96	1127
Absenteeism rate	0.03	0.03	0.03	0.03	(0.01)	0.75	(0.00)	0.72	(0.00)	0.46	974
Promotion rate	0.01	0.01	0.01	0.03	(0.01)	0.68	(0.00)	0.96	(0.00)	0.02	1127
Production Data											
Standard minute value	23.65	22.33	20.80	18.61	(3.47)	0.71	(1.62)	0.38	(0.88)	0.06	3083
Daily output	3750.59	3961.26	3401.75	4350.53	(678.45)	0.76	(300.33)	0.56	(167.49)	0.23	6308
Running days	9.31	11.89	8.99	10.18	(2.15)	0.23	(1.03)	0.88	(0.41)	0.48	4501
Efficiency	0.43	0.47	0.44	0.44	(0.05)	0.46	(0.02)	0.80	(0.01)	0.87	2903
Alteration rate	0.07	0.07	0.09	0.06	(0.02)	0.78	(0.01)	0.21	(0.00)	0.30	5144
Reject rate	0.00	0.00	0.00	0.00	(0.00)	0.90	(0.00)	0.63	(0.00)	0.87	5144

Notes: Adjacent lines are lines next to treatment lines. Non-survey lines are all lines in the factory which are neither control, treatment nor adjacent lines, i.e. all lines on which no surveys had been conducted. In some factories, some of the other lines are located on floors without any control, treatment or adjacent lines. (C-T), (C-A) and (C-Ns) indicate the difference of average values of control lines from those of treated lines, adjacent lines or non-survey lines respectively, with se and p showing the standard deviation and p-value of the difference. The data is at the line level and includes all observations within two months prior to the month of the training start. The F-test statistic of the omnibus balance test for control and treated lines using salary data is 0.50. The F-test statistic of the omnibus balance test for control and treated lines using production data is 1.37.

2.5 Impact on worker outcomes

2.5.1 Main effects

We begin by examining the effect of the training on management practices. We measure practices with responses from surveys of both line operators and line-level managers. We might expect line-level managers to overstate the extent to which they are carrying out practices, especially those directly taught in the training. While managers may carry out some practices without workers being aware they are doing so, the practices involving how workers ask for leave or conducting skills assessments cannot be implemented without workers being aware of them. Recall that the surveys are carried out on treatment and control lines, as well as on the lines adjacent to the treatment line. We use combinations of the lines to measure treatment effects and spillover effects, as described later in this section.

The survey asked workers whether several practices taught at the training were implemented on their line. The HR training modules, for example, introduced several practices related to the induction of new workers into the factory. The goal of these practices is to reduce the very high exit rates within the first few months after a worker joins the factory by making new workers feel more like they are part of the team.[5] In addition to induction, the training encouraged factories to make it easier for workers to ask for formal leave. In most factories, workers are able to take one day of leave per month without losing their attendance bonus, so long as they arrange the leave in advance. Often, arranging leave requires that the worker leave her machine to obtain signatures from the line chief, the HR officer and perhaps others. Given the sequential nature of production on the line, this process can cause a production bottleneck. Facing intense pressure to catch up when she returns to her station, the worker may instead choose instead simply to be absent without notice. This is costly both to the worker, who loses some part of her attendance bonus, and the factory, who must adjust workers to fill in

[5] A key practice is the assignment of a "buddy" for each new worker. The buddy is a line operator with substantial tenure at the factory who can answer the questions a new worker might have about practices and expectations at the factory. A separate induction practice taught in the training is having the Welfare Officer responsible for the line where the new worker is assigned meet with each new worker on the first day of work, to provide information about safety and other procedures at the factory.

the gap on the production line after production has already started for the day. The training taught the practice of "one-step leave", where the worker asks only her supervisor for permission to be absent, and the supervisor then obtains the required signatures. This both prevents bottlenecks and makes it more likely that workers will give notice when they need to be absent. Both the worker and the supervisor should be aware if this procedure has been implemented following the training.

We take a similar approach for practices related more directly to production. For example, the training encourages the factory to make regular skills assessments of workers. Sewing skills are usually the largest component of promotion decisions, so regular assessments helps to ensure that promotion decisions are viewed by workers as being fair. Skills assessments also provide line managers with current information about the capabilities of each operator, which is useful in assigning workers to specific tasks. A second practice stressed in the training is the need to provide continuous training to machine operators on the line. Section 2.B.3 in the Appendix shows the complete list of practices, and the list of questions we use to determine whether or not these practices are being followed on the line. Finally, we use survey responses related to worker attitudes to create measures of worker wellbeing and communication between workers and managers. The specific questions on which these measures are based are also shown in Section 2.B.3.

Empirical specifications: We begin by combining the several measures related to practices for management of new and existing production workers into two indices, one measuring adoption of HR practices, and the second adoption of production practices. We create the indices using the method suggested by Anderson (2008). We then estimate the following equation for each of these two indices:

$$I_{l,f,t} = \beta_0 + \beta_1 EL + \beta_2 T_{l,f} x ML + \beta_3 T_{l,f} x EL + \delta_f + \epsilon_{f,t} \tag{2.1}$$

where $I_{l,f,t}$ is the index associated with a set of practices which aggregates survey responses from all operators or line supervisors/line chiefs working on one particular line l in factory f in month t. $T_{l,f}$ is a treatment indicator equal to one if the survey respondents work on a treatment line. ML and EL are indicator variables indicating the responses are from midline and endline surveys, respectively.

$T_{l,f}xML$ and $T_{l,f}xEL$ are indicator variables equal to one if both the treatment indicator $T_{l,f}$ and ML, resp. $T_{l,f}$ and EL are equal to one. We also include factory fixed effects δ_f. The error term $\epsilon_{f,t}$ is clustered at the line level.

As many managers are not assigned to specific lines and share responsibility for several lines or even the whole factory, the specification for the manager surveys does not contain a treatment indicator variable. It measures a simple change in practices across time (between midline and endline)[6]:

$$I_{l,f,t} = \beta_0 + \beta_1 EL + \delta_f + \epsilon_{f,t} \tag{2.2}$$

Results and possible mechanisms: Results are depicted in Table 2.5, columns 1 (HR) and 6 (production) for line operators, columns 3 and 8 for line supervisors/line chiefs, and columns 5 and 10 for managers. We see that treatment had a positive but insignificant effect on HR practices, measured by responses from line operators and managers. In column 2, we add adjacent lines as controls, finding that the effects on HR practices are significant, though those on production-related practices remain insignificant. Adding the adjacent lines increases statistical power. However, the assignment to treatment and control is no longer random, as the adjacent lines were not among the three lines chosen by the factories as potential pilot lines. Moreover, if practices spillover to the adjacent lines, then the measured treatment effect will be biased downward. Table 2.6 shows the results for indices of worker wellbeing and communication. Worker wellbeing of operators is slightly lower on treatment lines at the time of the midline survey, though the effect is not significant, but is marginally significantly higher at the time of the endline survey. We do not see any improvement in the wellbeing of line supervisors/line chiefs and managers, though the direct focus of the programme was on production workers in any case. The clearest and largest effects of the training appear to be on communication between workers and managers. The communication index based on the line operator surveys shows a large and highly significant difference between treatment and control groups.

[6]Managers were not surveyed at baseline.

TABLE 2.5. HR AND PR INDEX: ALL CONTROL LINES

	HR Index					PR Index				
	(1) LO	(2) LO	(3) LSLC	(4) LSLC	(5) MG	(6) LO	(7) LO	(8) LSLC	(9) LSLC	(10) MG
Endline	-0.063	-0.15	-0.086	-0.23		-0.033	-0.0025	-0.12	-0.15	
	(0.30)	(0.28)	(0.31)	(0.30)		(0.27)	(0.24)	(0.27)	(0.24)	
MidlineXTreated	0.46	0.52*	0.20	0.035		0.16	0.15	0.33	0.37	
	(0.31)	(0.28)	(0.28)	(0.25)		(0.30)	(0.29)	(0.30)	(0.30)	
EndlineXTreated	0.30	0.44	0.0041	0.011		0.24	0.15	0.18	0.20	
	(0.31)	(0.32)	(0.31)	(0.31)		(0.30)	(0.28)	(0.36)	(0.37)	
MidlineXAdjacent		0.33		-0.51**			-0.36		-0.50*	
		(0.28)		(0.23)			(0.28)		(0.27)	
EndlineXAdjacent		0.25		-0.067			-0.090		-0.36	
		(0.27)		(0.28)			(0.31)		(0.26)	
Treated					0.24**					-0.32**
					(0.11)					(0.12)
Constant	0.15	0.18	0.12	0.25**	-0.11**	0.046	0.011	0.16	0.12	0.15**
	(0.18)	(0.17)	(0.15)	(0.13)	(0.053)	(0.14)	(0.18)	(0.20)	(0.17)	(0.056)
Observations	126	209	128	219	247	126	209	128	219	247
Line clusters	63	106	65	116	25	63	106	65	116	25
p: MLxT=MLxA		0.53		0.055			0.084		0.0086	
p: ELxT=ELxA		0.50		0.77			0.43		0.087	

Notes: Table 2.5 shows the main treatment effects on the HR index and PR index. All indices are based on Anderson (2008). Index components are different for line operators (LO), line supervisors/line chiefs (LSLC) and managers (MG) (see Section 2.B.3 in the Appendix). Midline and Endline are indicator variables equal to one if responses are from midline and endline surveys, respectively. Adjacent lines are lines next to treatment lines. MidlineXTreated and EndlineXTreated are indicator variables equal to one if both the treatment indicator and the respective survey indicator are equal to one. MidlineXAdjacent and EndlineXAdjacent are indicator variables equal to one if the adjacent line indicator and the respective survey indicator are equal to one. Regressions include factory fixed effects and the error term is clustered at the line level. p-values are shown at the bottom of the table for testing the equality of coefficients on treatment and adjacent lines at midline and endline (MLxT=MLxA; ELxT=ELxA). Standard errors are shown in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Collectively, the results on Table 2.5 and 2.6 show that workers on the production floor benefitted from the programme, in particular with respect to HR practices, wellbeing and communications. Production practices, on the other hand, do not appear to have changed following the training.

We show an overview of disaggregated results for each component of the indices in Figure 2.1, which summarizes results from the Appendix Tables 2.D.5, 2.D.6 and 2.D.7. Columns 1 and 2 show that the buddy system, one-step leave policy, and communication approaches seem to have worked particularly well. All of these practices change the way workers interact on the production floor. In contrast, the HR-related practices related to promotion policies or job descriptions,

TABLE 2.6. WELLBEING AND COMMUNICATION INDEX: ALL CONTROL LINES

	Wellbeing Index					Communication Index				
	(1) LO	(2) LO	(3) LSLC	(4) LSLC	(5) MG	(6) LO	(7) LO	(8) LSLC	(9) LSLC	(10) MG
Endline	0.58**	0.53**	0.46	0.50		0.54**	0.54**	-0.024	-0.10	
	(0.23)	(0.21)	(0.30)	(0.31)		(0.23)	(0.21)	(0.30)	(0.26)	
MidlineXTreated	-0.23	-0.26	0.13	0.15		1.07***	0.98***	0.045	-0.038	
	(0.31)	(0.29)	(0.26)	(0.30)		(0.24)	(0.23)	(0.26)	(0.21)	
EndlineXTreated	0.41*	0.47*	0.20	0.20		1.13***	0.98***	0.29	0.24	
	(0.24)	(0.24)	(0.26)	(0.27)		(0.28)	(0.27)	(0.36)	(0.36)	
MidlineXAdjacent		0.022		-0.094			1.22***		-0.54**	
		(0.29)		(0.30)			(0.26)		(0.22)	
EndlineXAdjacent		0.31		-0.0039			0.52**		0.23	
		(0.27)		(0.31)			(0.26)		(0.27)	
Treated					-0.039					0.14
					(0.16)					(0.16)
Constant	-0.34*	-0.31	-0.16	-0.20	0.018	-0.14	-0.10	0.20	0.27*	-0.065
	(0.19)	(0.19)	(0.19)	(0.22)	(0.073)	(0.16)	(0.16)	(0.17)	(0.15)	(0.072)
Observations	126	209	128	219	247	128	209	128	219	247
Line clusters	63	106	65	116	25	65	106	65	116	25
p: MLxT=MLxA		0.35		0.37			0.39		0.016	
p: ELxT=ELxA		0.59		0.48			0.14		0.96	

Notes: Table 2.6 shows the main treatment effects on the Wellbeing and Communication index. All indices are based on Anderson (2008). Index components are different for line operators (LO), line supervisors/line chiefs (LSLC) and managers (MG) (see Section 2.B.3 in the Appendix). p-values are shown for testing the equality of coefficients on treatment and adjacent lines at midline and endline (MLxT=MLxA; ELxT=ELxA). Standard errors are shown in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

and most production-related practices and tools, skills tests, skills matrices, etc. do not seem to have been implemented, or are even worse than on control lines. From the perspective of line supervisors/line chiefs (columns 3 and 4), only quality practices seem to have been implemented. One core part of quality practices is the introduction of zero-defect operators who are responsible for checking the quality of certain processes. This practice might have also changed how operators interact with each other on the production floor. Further, supervisors/line chiefs feel that workers are more involved, measured by the extent to which workers take over responsibility for changes on the line and suggest ideas. From the perspective of managers (column 5), the buddy system is fully running at endline. As the coefficients measure changes between midline and endline, the negative coefficient on quality practices in the manager survey does not imply a halt of those practices, but rather indicates that no additional quality practices have been im-

	Surveyed group				
	Line operators		Line supervisors and line chiefs		Managers
	Independent variables				
Dependent variables	(1) MidlineX Treated	(2) EndlineX Treated	(3) MidlineX Treated	(4) EndlineX Treated	(5) Endline
HR Index	0.52**	0.44	0.035	0.011	0.24**
Induction system	0.36	-0.020			
Buddy system	1.32***	0.51**			0.64***
One-step leave policy	0.11	0.54**	-0.074	0.17	0.075
Worker representative	-0.10	0.38			
Promotion policy	-0.27	-0.16			-0.16
Job descriptions			0.25	-0.16	
Exit interviews					-0.023
Attendance bonus	0.0035	-0.042	-0.070	-0.011	-0.13
PR Index	0.15	0.15	0.37	0.20	-0.32*
Up-skilling	-0.035	0.33	0.23	0.0031	-0.013
Skills test	0.19	-0.034	0.0086	0.40	0.20
Quality practices			0.72***	0.050	-0.68***
Production-related practices			0.26	-0.30	-0.20
Production-related tools			0.53	0.49	
Wellbeing Index	-0.26	0.47*	0.15	0.20	-0.039
Feelings of surveyed group	-0.20	0.46**	-0.27	-0.27	-0.072
Operators' feelings					0.0075
Involvement of workers	-0.11	0.098	0.38*	0.43*	
Communication Index	0.98***	0.98***	-0.038	0.24	0.14
Communication practices	0.86***	1.32***	0.29	0.16	0.14
Knowledge about LO/LS/LC	0.068***	0.033	-0.26	0.039	
Communication with Industrial Engineer			0.12	0.30	

Figure 2.1: Illustration of effects for indices and sub-indices

Notes: Figure 2.1 illustrates the OLS estimates of treatment effects on the four indices HR, PR, Wellbeing and Communication, as well as their sub-indices including all control lines on both treatment and control floors. Outcome variables are listed on the left. Column (1) and (3) resp. (2) and (4) report the estimated treatment effects at midline resp. endline for line operators and line supervisors/chiefs from a regression of the outcome variables on the treatment indicator interacted with the survey time, including separate interaction terms for adjacent lines. Column (5) reports the estimated effects over time for managers from a regression of the outcome variables on the endline indicator. Blank cells indicate that the sub-index is non-existant, i.e. the surveyed group was not asked any questions about that particular topic. All indices are constructed using Anderson (2008) variance-covariance weighted index. Comprehensive results of this illustration, including standard errors, line clusters and p-values for tests on the similarity of coefficients, are shown for overall indices in Tables 2.5 and 2.6 (columns 2, 4, 5, 7, 9 and 10 each), and for sub-indices in Tables 2.D.5 (line operators), 2.D.6 (line supervisors/chiefs) and 2.D.7 (managers). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

plemented between midline and endline.[7] We provide qualitative support of the disaggregated results, by showing managers' perceptions about the most impor-

[7] 91% of the managers indicate at midline, that the factory has implemented a new quality system in the past six months. At Endline, this percentage is only 71%. As the time between midline and endline surveys is larger than six months, a negative coefficient can simply suggest that no additional quality systems have been implemented at endline, and the already implemented quality systems at midline are still running.

tant changes in policies or processes. Even managers from the production, IE and quality department consider HR-related practices, such as the buddy system and the one-step leave policy, among the most important changes (see left graph of Figure 2.D.19). Interestingly, they consider teamwork as one of the most important changes in their own job (see left graph of Figure 2.D.20), which also relates to the interaction patterns on the production floor.

The surveys clearly indicate that practices aiming at directly changing the interaction of people seem to have been better implemented than practices aiming at changing technical aspects and the production process itself. To explain this result, we draw upon a detailed evaluation of each practice (blind to the outcomes) by industry experts as described in Appendix Section 2.A.3. Based on set-up costs, operating/maintenance costs and the speed and precision of feedback, all practices are ranked with the highest number being the lowest cost practice to implement. Practices which worked well are among the most highly ranked practices, e.g. the buddy system and one-step leave policy are tied with the second highest ranking. Looking at the criteria, one possible explanation is that the technical practices require more cross-departmental communication to be approved and implemented. That is, several departments need to be involved in implementing and using the practices. For example, the skills assessments are carried out by industrial engineers, but used by industrial engineers and production to assign workers to tasks and by industrial engineers and HR to determine promotions. Many practices affecting the interaction of workers on the production floor require a less intensive exchange of information between different departments (e.g. only HR and Production department). Implementation of one-step leave, for example, requires only that the production department change its practice. As described in Section 2.2.2, decisions requiring several departments are often delayed or fail due to the hierarchical nature of the management structure. Hence, practices including several departments might have been more difficult to implement. There are other possible explanations as well, for example, that the production practices are factory-specific and therefore require a high degree of customisation. HR practices may have had lower setup costs for other reasons, or there may have been quicker and more easily measured feedback about whether or not practices worked. Fur-

thermore, we can conclude that not all practices were complementary; successfully adopting some practices by themselves led to positive effects on worker outcomes.

2.5.2 Spillover effects

In this section, we aim to analyse whether adjacent lines and/or control lines adopted the same practices implemented on treatment lines, possibly leading to an underestimation of the results in the analysis in Section 2.5.1. We focus on practices/approaches that appear to have worked well on treatment lines from the workers' point of view, i.e. the HR, wellbeing and communication index for line operators as described above.

Empirical specification: The sample for regressions measuring spillover effects includes all lines in the factory. The adjacent lines are not randomly assigned, which should be kept in mind when interpreting the results. The specification is as follows:

$$I_{l,f,t} = \beta_0 + \beta_1 EL + \beta_2 T_{l,f} xML + \beta_3 T_{l,f} xEL + \beta_4 A_{l,f} xML + \beta_5 A_{l,f} xEL + \delta_f + \epsilon_{f,t} \quad (2.3)$$

where $A_{l,f}$ is an indicator equal to one if the survey respondents work on an adjacent line. $A_{l,f}xML$ and $A_{l,f}xEL$ are indicator variables equal to one if both the adjacent indicator if $A_{l,f}$ and ML, resp. $A_{l,f}$ and EL are equal to one. All other parameters are specified as in Equation 2.1. Second, we aim to measure spillovers from treatment to control lines, by estimating Equation 2.1 and 2.3 with different restrictions on the control group. We use only control lines on the same floor as treatment lines (see column 3, Figure 2.C.13), or only control lines on a different floor than treatment lines (see column 4, Figure 2.C.13). Stronger treatment effects in the regression comparing treated lines with control lines on other floors ("Diff") would be an indication of spillover effects, indicating perhaps that control lines on the same floor started to adopt the same practices as on treatment lines.

Results and possible interpretation: We measure spillovers to the lines adjacent to the treatment lines by including the adjacent lines in the sample and adding interaction terms MidlineXAdjacent and EndlineXAdjacent. The results are shown

in columns 2, 4, 7 and 9 in Tables 2.5 (HR and production practices) and 2.6 (well-being and communications). We observe positive significant spillovers to adjacent lines only with respect to communication practices (see Table 2.6). The coefficients on MidlineXAdjacent and EndlineXAdjacent for the operator's HR and wellbeing index are positive but not significant.

The spillover regressions in Tables 2.5 and 2.6 include all of the control lines. Recall that in two-thirds of the factories, we have a control line on a different floor than the treatment line. Both because communication among managers across floors is much less frequent than communication within floors, and because some practices might be implemented at the floor level, we arguably obtain cleaner results if we exclude control lines on the same floor. The results from regressions making this exclusion are shown in columns 4, 8 and 12 of Table 2.7 (labeled "Diff"). We continue to find significant spillover effects to adjacent lines for the communication index, but now also find marginally significant effects for the HR index using the endline data. Coefficients of adjacent lines for the wellbeing index go in the same direction as those of treatment lines (first negative, then positive), but stay insignificant.

Spillovers to control lines on the same floor are only observed for the HR index. Effects on wellbeing seem to have been restricted to treatment lines, and effects on communication to treatment and adjacent lines without any spillovers to control lines. These results can be seen by comparing regressions with control lines on the same floor to regressions with control lines on a different floor. Effects for the HR index disappear in the former (column 1 and 3, Table 2.7), but are even stronger in the latter (column 2 and 4, Table 2.7). Effects for the communication index, however, stay significant in both specifications (columns 9 to 12, Table 2.7).

To summarize, we find evidence of spillovers of HR-related practices to adjacent and control lines on the same floor, but spillover of communication practices only to adjacent lines. We find no spillovers on worker wellbeing. One possible explanation for this pattern is that HR-related practices are implemented at the floor level (e.g., induction and buddies, which are often implemented before the worker is assigned to a fixed line) or are very well-defined, easy-to-deploy practices/tools that can be copied without intensive coordination processes (e.g., one-step leave).

TABLE 2.7. SPILLOVER ANALYSIS: LINE OPERATORS ONLY

	HR Index				Wellbeing Index					Communication Index		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Same	Diff	Same	Diff	Same	Diff	Same	Diff	Same	Diff	Same	Diff
Endline	0.10	-0.14	0.018	-0.31	0.55	0.59**	0.42	0.51*	0.99**	0.34	1.00**	0.32
	(0.44)	(0.41)	(0.41)	(0.35)	(0.45)	(0.28)	(0.39)	(0.26)	(0.48)	(0.25)	(0.44)	(0.22)
MidlineXTreated	0.23	0.56	0.094	0.70**	0.032	-0.34	-0.31	-0.28	1.35***	1.01***	1.22***	0.85***
	(0.44)	(0.36)	(0.40)	(0.32)	(0.32)	(0.39)	(0.33)	(0.35)	(0.38)	(0.27)	(0.34)	(0.25)
EndlineXTreated	-0.089	0.47	-0.16	0.78**	0.70**	0.29	0.53*	0.47	0.95***	1.27***	0.77**	1.07***
	(0.45)	(0.35)	(0.44)	(0.35)	(0.33)	(0.31)	(0.32)	(0.30)	(0.35)	(0.35)	(0.38)	(0.30)
MidlineXAdjacent			-0.10	0.50*			-0.022	-0.013			1.47***	1.09***
			(0.41)	(0.30)			(0.33)	(0.35)			(0.38)	(0.27)
EndlineXAdjacent			-0.36	0.58*			0.39	0.29			0.30	0.61**
			(0.42)	(0.30)			(0.33)	(0.33)			(0.40)	(0.26)
Treated												
Constant	0.37	0.047	0.57*	0.036	-0.53**	-0.28	-0.27	-0.29	-0.49*	-0.0035	-0.38	0.058
	(0.34)	(0.24)	(0.32)	(0.21)	(0.21)	(0.26)	(0.25)	(0.26)	(0.29)	(0.19)	(0.29)	(0.18)
Observations	76	104	159	187	76	104	159	187	78	106	159	187
Line clusters	38	52	81	95	38	52	81	95	40	54	81	95
p: MLxT=MLxA			0.51	0.52			0.33	0.38			0.42	0.40
p: ELxT=ELxA			0.49	0.49			0.64	0.57			0.13	0.15

Notes: Table 2.7 shows regression results for HR, Wellbeing and Communication index based on survey data from line operators. "Same" indicates that only control lines on the same floor as treatment lines are included, whereas "Diff" indicates that only control lines on a different floor than treatment lines are included. All indices are based on Anderson (2008). p-values at the bottom are shown for testing the equality of coefficients on treatment and adjacent lines at midline and endline (MLxT=MLxA; ELxT=ELxA). Standard errors are shown in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Communication practices, on the other hand, and good working relationships between supervisors and workers are more of an "art" and a soft skill inherent to those line managers of treatment lines who attended the training.

2.6 Impact on firm outcomes

2.6.1 Main effects

We see changes in practices related to HR, and particularly those related to induction of new workers. We also see increased communication between workers and managers. Do either of these translate into changes in performance? We use administrative records from the factory to examine two sets of outcomes. First, we examine outcomes that we view as following directly from HR practices – exit of newly-hired workers, labour turnover among production workers more broadly, and absenteeism rates. Second, we look for effects of the training on productive efficiency and product quality.

Empirical specification: To analyse effects on firm outcomes, we define four post-training periods, corresponding with the four periods when the training modules are delivered (see Figure 2.A.2 in the Appendix). Each period corresponds to the delivery of one HR module and one production module. We define each post-training period as starting with the date following the delivery of one of the four training modules and ending with the start of the delivery of the subsequent training module. We define the fourth period as the three months after the fourth training module. Recall that the training is spread across six months. The regressions therefore show the pattern of the outcome variables over roughly nine months following the start of the training. The treatment effects might appear after any of these training periods. We estimate the following ANCOVA specification as illustrated in Figure 2.C.14 in the Appendix Section 2.C.3:

$$Y_{l,f,t} = \beta_0 + \beta_1 Ybase_{l,f} + \beta_2 Post2_{l,f} + \beta_3 Post3_{l,f} + \beta_4 Post4_{l,f} + \beta_5 TxPost1_{l,f} + \beta_6 TxPost2_{l,f} + \beta_7 TxPost3_{l,f} + \beta_8 TxPost4_{l,f} + \gamma_t + \delta_f + \epsilon_{f,t} \tag{2.4}$$

where $Y_{l,f,t}$ is the outcome variable of interest; l indexes the line in factory f in month t. $Ybase_{l,f}$ is the baseline value of the dependent variable (average over a period of three months before training start) and $T_{l,f}$ an indicator variable equal to one for the treatment line and equal to zero for the two control lines. $PostX_{l,f}$ is an indicator variable for the period between training X and X+1. $T_{l,f}xPostX$ are indicator variables equal to one if both $T_{l,f}$ and $PostX_{l,f}$ are equal to one. We also include monthly time dummies γ_t to control for seasonal variations, and factory fixed effects δ_f to control for systematic differences across factories. The error term $\epsilon_{l,f,t}$ is clustered at the line level. All indices are based on Kling et al. (2007).

Results and interpretation: Table 2.8 presents the main results of the effect of the training programme. We find a significantly positive effect after the first and second training modules for HR-related outcome variables, but not for production-related variables. There is a non-trivial increase in efficiency, though the effect is not statistically significant. Looking at the individual outcome variables of the HR index, we observe that the effect on the HR index is mainly driven by a sharp decrease in exit rates of newly hired workers. These exit rates fall by 16 percent,

TABLE 2.8. OVERALL EFFECTS: ALL CONTROL LINES
(COLUMNS 1 AND 2 IN SUMMARY TABLE 2.9)

	Index		**HR Variables**				**PR Variables**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	HR	**PR**	Exit	Turnover	Absenteeism	Promotion	Efficiency	Alterations	Rejects
TreatedXPost1	0.436*	0.041	-0.076	-0.030*	-0.002	0.012	0.037	0.000	0.000
	(0.239)	(0.187)	(0.061)	(0.017)	(0.008)	(0.014)	(0.023)	(0.010)	(0.001)
			{0.497}	{0.449}	{0.793}	{0.719}	{0.525}	{1.000}	{1.000}
TreatedXPost2	0.444*	0.001	-0.162**	-0.035	0.001	0.002	0.058	0.006	0.000
	(0.222)	(0.133)	(0.068)	(0.022)	(0.006)	(0.017)	(0.037)	(0.009)	(0.000)
			{0.088}	{0.204}	{0.836}	{0.836}	{0.641}	{1.000}	{1.000}
TreatedXPost3	-0.044	0.012	0.181	0.001	0.000	0.029	0.021	0.001	-0.000
	(0.407)	(0.141)	(0.169)	(0.016)	(0.008)	(0.022)	(0.042)	(0.010)	(0.000)
			{1.000}	{1.000}	{1.000}	{1.000}	{1.000}	{1.000}	{1.000}
TreatedXPost4	-0.133	0.093	0.000	0.012	0.001	-0.006	0.019	0.000	-0.001**
	(0.156)	(0.108)	(0.039)	(0.014)	(0.006)	(0.005)	(0.026)	(0.008)	(0.000)
			{1.000}	{1.000}	{1.000}	{1.000}	{0.928}	{1.000}	{0.065}
Observations	289	2069	289	289	246	289	846	2068	2068
Clusters	51	68	51	51	43	51	35	68	68
R^2	0.34	0.32	0.28	0.28	0.52	0.23	0.54	0.28	0.35
Control mean	0.02	-0.05	0.23	0.06	0.04	0.01	0.38	0.06	0.00
Factory FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes: Table 2.8 shows treatment effects on HR and PR indices in column one and two, which refer to columns 1 and 2 in the Summary Table 2.9, and individual index components (exit, turnover, absenteeism and promotion rates for HR index, and efficiency, alteration rates and reject rates for PR index). Regressions include time fixed effects to control for seasonal variations. We use weekly time dummies for the PR index and corresponding index components, and monthly time dummies for the HR index and corresponding index components. Factory fixed effects are also included to control for systematic differences across factories. The error term is clustered at the line level. False discovery rate-adjusted p-values, also known as q-values, for HR-related and PR-related outcome families were used to adjust for multiple outcome variables. They are shown in curly brackets. We use the 'sharpened q value' approach based on the Benjamini-Hochberg procedure (Benjamini et al., 2006). For simplicity, the regression table does not report coefficients on the $PostX$ variables for $X = (1, 4)$, and the baseline value of the dependent variables.

and the effect remains significant even when using sharpened q-values to account for multiple hypothesis testing. Thus, the administrative data reinforce the results reported in Section 2.5 above, in particular the strong effect on outcome variables driven by HR-related practices. Interestingly, Table 2.8 also shows that the overall effect on the outcome index vanishes after the third and fourth training date.

2.6.2 Spillover effects

In this section, we analyse spillovers from treatment lines to adjacent and/or control lines, possibly leading to underestimated results in Section 2.6.1. Furthermore, we aim to find out whether the disappearance of effects on the HR index after the

third and fourth training modules can be explained by the fact that treatment lines stop doing the new practices or by the fact that control lines adopted new practices as well, i.e. spillover effects.

Empirical specification: The factory data allows us to undertake a more detailed spillover analysis than the survey data, as it also contains outcome data for lines where we did not survey workers. Factories with more than two floors may have purposely chosen lines on floors that are different from the other floors in the factory. However, since we randomly selected the treatment line from the lines nominated by the factory, we should expect that, on average, the lines on floors with a treatment line will be identical to the lines on floors with at least one control line.

Recall that we define factories as A-type if they have a control line on a different floor than the treatment line, as B-type if they have control lines on both the treatment and the control floor, and as C-type if they have control lines only on the same floor as the treatment line. We use eight specifications to measure spillovers (see Figure 2.C.13 in the Appendix) as follows:

(i) "All C-lines" (ABC-types): Treatment lines are compared with control lines in all sample factories of type A, B and C. This is the basic specification as described in Section 2.6.1.

(ii) "All C-lines" (AB-type): We exclude all factories which only have control lines on the same floor as treatment lines (factories of type C). As the following specifications to measure spillovers only include A and B factories, this specification serves as a better comparison than the specification with A, B and C factories in Section 2.6.1.

(iii) "Diff C-lines" (AB-types): Effects in specification ii might be underestimated due to control lines on the same floor as treatment lines. To analyse whether results are actually confounded, we compare treatment lines only to control lines on a different floor than the treatment line.

(iv) "Adj-lines" (ABC-type): We include adjacent lines in the basic specification i to analyse spillover effects from treatment to adjacent lines which are located next to treatment lines on the same floor.

(v) "Adj-lines" (AB-type): We use specification iv as a starting point, but only include control lines on different floors than treatment lines (without factories of type C). We expect results without control lines on the treatment floor, to which practices might have easily spilled over, to be less confounded than results of specification iv with control lines on the treatment floor.

(vi) "Adj+Other" (AB-types): We include adjacent and non-survey/other lines on the treatment floor and compare those to all lines on the control floor. This enables us to analyse spillover effects from treatment lines to adjacent and/or other lines on the treatment floor.

(vii) "C-floor" (AB-types): We compare all lines on treatment floors against all lines on control floors. This allows to detect treatment effects under the circumstance that practices spilled over to the whole treatment floor.

(viii) "All floors" (AB-types): We compare all lines on treatment floors against all lines in the factory. This specification is similar to specification iv, but less robust due to the fact that other floors without any control lines might be different to treatment floors.

Results and interpretation: As there is no significant effect on the PR index, we only focus on the HR index in the following paragraph. All results refer to Table 2.9. A key takeaway from the analysis is that although the results in the short-run generally hold up to controlling for spillovers, we find no significant effects for any outcome in period 4, even when we compare treatment lines to lines on other floors. While it is possible that the effects spill over even to other floors, the analysis of practices in the previous section indicates that the differences in practices remain even after the endline survey, which is roughly period 4 in the administrative data.

Excluding control lines on the same floor as treatment lines leads to positive effects on the HR index after the first, second and third training, which then disappears after the fourth training (see column 5). If control lines on the same floor are included in the analysis, the coefficients on the treatment indicator after the

TABLE 2.9. SUMMARY TABLE OF OVERALL EFFECTS USING DIFFERENT FACTORY TYPES AND LINES

	All C-lines (ABC)		All C-lines		Diff C-lines		Adj-lines (ABC)		Adj-lines		Adj+Other		C-floor		All floors	
	(1) HR	(2) PR	(3) HR	(4) PR	(5) HR	(6) PR	(7) HR	(8) PR	(9) HR	(10) PR	(11) HR	(12) PR	(13) HR	(14) PR	(15) HR	(16) PR
TreatedXPost1	0.44*	0.041	0.21	0.0020	0.28	0.016	0.45*	0.037	0.29	0.016	0.20	0.045	0.11	0.12	0.045	0.044
	(0.24)	(0.19)	(0.21)	(0.23)	(0.22)	(0.24)	(0.23)	(0.19)	(0.21)	(0.23)	(0.13)	(0.22)	(0.095)	(0.10)	(0.091)	(0.093)
TreatedXPost2	0.44*	0.0011	0.45	-0.0042	0.49	-0.057	0.46**	-0.00032	0.51	-0.050	0.47	-0.12	0.12	-0.023	0.27	0.0044
	(0.22)	(0.13)	(0.44)	(0.15)	(0.45)	(0.15)	(0.21)	(0.13)	(0.42)	(0.15)	(0.35)	(0.13)	(0.20)	(0.060)	(0.18)	(0.055)
TreatedXPost3	-0.044	0.012	0.40	0.0047	0.45	-0.0089	0.018	0.010	0.46	-0.0043	0.60	-0.053	0.69**	-0.035	0.81***	0.016
	(0.41)	(0.14)	(0.49)	(0.16)	(0.51)	(0.17)	(0.40)	(0.14)	(0.52)	(0.17)	(0.50)	(0.15)	(0.30)	(0.072)	(0.30)	(0.067)
TreatedXPost4	-0.13	0.093	-0.20	0.088	-0.17	0.070	-0.12	0.093	-0.15	0.084	-0.054	0.076	0.11	0.070	0.12	0.052
	(0.16)	(0.11)	(0.23)	(0.12)	(0.23)	(0.12)	(0.15)	(0.11)	(0.22)	(0.12)	(0.19)	(0.10)	(0.074)	(0.051)	(0.073)	(0.049)
AdjacentXPost1							0.16	0.062	-0.022	0.011	-0.13	0.055				
							(0.19)	(0.15)	(0.22)	(0.19)	(0.16)	(0.16)				
AdjacentXPost2							-0.014	0.096	0.39	0.046	0.34	-0.054				
							(0.24)	(0.10)	(0.44)	(0.11)	(0.38)	(0.084)				
AdjacentXPost3							-0.35	-0.070	0.33	-0.00082	0.41	-0.049				
							(0.56)	(0.13)	(0.53)	(0.15)	(0.48)	(0.13)				
AdjacentXPost4							0.0061	0.059	0.12	0.036	0.20*	0.042				
							(0.11)	(0.094)	(0.14)	(0.11)	(0.10)	(0.084)				
OtherXPost1											0.14	0.18				
											(0.11)	(0.12)				
OtherXPost2											-0.067	0.0054				
											(0.22)	(0.067)				
OtherXPost3											0.86***	-0.042				
											(0.33)	(0.075)				
OtherXPost4											0.13	0.084				
											(0.085)	(0.057)				
Observations	289	2069	200	1631	188	1511	469	3397	297	2486	1080	8107	1080	8107	1245	9689
R^2	0.34	0.32	0.34	0.33	0.32	0.32	0.29	0.31	0.31	0.32	0.26	0.31	0.25	0.31	0.21	0.29
Line clusters	51	68	35	52	33	48	83	112	52	79	193	257	193	257	224	306
Control Mean	0.024	-0.047	0.048	0.022	0.073	0.017	0.042	-0.047	0.17	-0.0066	0.077	0.045	0.0095	0.0011	0.018	-0.0011
Factory Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Time Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes: Table 2.9 shows the main treatment effects on firm-related outcome variables, i.e. the HR and PR index. All indices are based on Kling et al. (2007). Columns 1 to 6 show regression results of treatment lines (T) compared to control lines, out of which columns 3 to 6 exclude factories with control lines only on the same floor as treatment lines. Columns 7 to 10 compare treatment lines (T) and adjacent lines (A) to control lines, out of which columns 9 and 10 exclude factories with control lines only on the same floor as treatment lines and only keep control lines on control floors. Columns 11 and 12 compare treatment lines (T), adjacent lines (A) and and other lines on the treatment floor (O) to all lines on the control floor. Columns 13 and 14 compare treatment floors to the control floors, and columns 15 and 16 against all lines on all other floors in the factory. Regression results for individual index components are displayed in Section 2.E. Standard errors are shown in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

first, second and third training are still positive, but smaller in magnitude, indicating that spillovers might have taken place to control lines on the same floor (see column 3). The coefficients are not significant, though this could owe to a lack of statistical power.

We can only see positive (but insignificant) coefficients after the second and third training for adjacent lines once control lines on the same floor as treatment lines are excluded (see column 7 without exclusion and column 9 with exclusion). This suggests that there may be spillover effects not only to adjacent, but also to control lines on the same floor.

We can confirm the spillover of new HR practices to control lines on the same floor as treatment lines, as indicated by a positive and significant coefficient after the third period (OtherXPost3 in column 11), and positive (but not significant) coefficient in the fourth period. The positive and significant coefficient only appears after the third training, possibly because a floor-wide spillover to *all other* lines on the floor might take some time. The spillover effect on adjacent lines seems to already after the second training with positive (but not significant) coefficients until the third training and significantly positive coefficients after.

If we compare all lines on the treatment floor to all lines on the control floor (column 13) and to all other lines in the factory (column 15), a positive and significant coefficient appears after the second training as well, which is in line with the results separated for adjacent and other lines in column 11. The effect seems to have decreased after the fourth training date, which might be due to the fact that some lines stopped doing certain practices or some practices were dropped. However, we cannot make the conclusion that all practices were dropped, as we do not see a decrease of effects on the HR, wellbeing or communication index in the survey data (see Section 2.5.1). Results on treatment lines are still significantly better at the "Endline"-date (used for survey data), which is even several months after the "Post4"-date (used for factory data).

Regressions with individual index components using the specifications above are shown in Appendix Section 2.E. For example, Table 2.E.16 shows that HR spillover effects to adjacent and other lines on the treatment floor are driven by

a decrease in exit rates, turnover and absenteeism. Efficiency seems to improve also on other lines on the same floor as treatment lines.

Robustness: Many factories do not update administrative records when workers are moved across production lines. Hence, the administrative data may not accurately reflect the line on which a worker is actually working. However, we know the line code on the day any worker is surveyed from the survey data. We can use the survey data to test the accuracy of the administrative records for each factory. We re-run the regressions from above with a subset of factories that have the most data at the floor level, and with another subset of factories that have the most accurate data at the line level. We determine the two subsets of factories by matching the line code/floor in the survey data with the line code/floor data in the factory administrative records for the month of the survey as an indication of the accuracy of the administrative records. As workers move more frequently across lines on the same floor, but only rarely to lines on a different floor, the subset with accurate floors only excludes five factories (with a match rate <75%), whereas the subset with accurate lines drops thirteen factories (with a match rate <60%). Running regressions with the subset of factories limits the statistical power, but shows similar patterns of results (see Appendix Tables 2.E.23 and 2.E.24 for the subset based on floors, and Tables 2.E.25 and 2.E.26 for the subset based on line codes).

2.7 Conclusion

We implemented a group-based management training programme aimed first at improving conditions for workers in 25 garment factories in Bangladesh. In addition to the particular focus on improving working conditions alongside productivity, the programme was distinctive in teaching relatively generic practices in groups of firms, as well as the joint implementation of all practices on one single line per factory to account for complementarity of practices. We find an overall positive effect of the training programme on worker and factory outcomes suggesting that outcomes can be improved not only by implementing factory-specific practices as in Bloom et al. (2013) or Iacovone et al. (2018), but also by adopting generic practices, with only help with "activations" at the factory level.

To examine effects on worker outcomes, we analysed detailed survey data collected from line operators, line supervisors/line chiefs and managers. We find positive and significant effects on HR practices, wellbeing and communication approaches for line operators. This indicates that workers on the production floor benefited from the programme. Practices aimed at changing the interaction of workers and managers worked particularly well (e.g. buddy system or one-step leave policy), whereas technical and production-related practices (e.g. job descriptions, or skills matrices) do not show significant effects. We posit that this might reflect the greater extent of cross-departmental communication required for implementation of these practices. Due to the hierarchical nature of the management structure, decisions involving several departments require complex coordination processes leading to delays or failures of decision-making processes and possibly to fewer adoptions of those practices that require more cross-departmental communication.

We also find that the practices that were adopted lead to hard outcomes at the factory level - a reduction in exit rates among newly-hired workers, for example. This suggests that at least some practices are effected when adopted on their own, and do not have high degrees of complementary. The adoption of practices changing the interaction of workers alone leads to improvements of outcome variables. The data collection was designed to detect spillovers to adjacent lines and lines on the same production floor. We find positive and significant spillover effects for HR practices to adjacent and control lines on the same floor. Effects on wellbeing seem to be more local and restricted to treatment lines only, while effects on communication affect both treatment and adjacent lines, but not other lines on the floor. We conjecture that this is due to the higher extent of soft skills and less concrete/easy-to-implement nature of communication practices and good working relationships between supervisors and workers. In addition, some of the HR practices were most likely implemented floor-wide from the beginning, for example, those related to induction of new workers.

To analyse effects on firm outcomes, we used detailed factory administrative data (salary and production data). We find positive and significant effects on HR-related outcomes, in particular driven by a decrease in exit rates (less workers leav-

ing the factory within two months of joining). Effects on PR-related outcomes are not significant: Efficiency is insignificantly higher on treatment lines than control lines. The disappearance of the positive effect on HR-related outcomes on treatment lines after the third training module can be explained by spillover effects of practices to other lines on the same floor as treatment lines. In particular, we can observe a decrease in exit rates, turnover and absenteeism on treatment floors. Spillover effects on PR-related outcome variables are weaker. Efficiency seems to improve on other lines on the same floor as treatment lines, whereas alteration rates are as high on treatment as on control lines. All effects on both HR and production outcomes seem to vanish after the last training, possibly due to the discontinuation of certain practices. As survey indices are still positively significant at endline (which is several months after the fourth and last training), it is likely that at least some practices have been kept.

Acknowledgements We would like to thank the local consulting team and the participating factories for their support and cooperation. We would like to thank the field managers and enumerator team at the Innovations for Poverty (IPA) office in Bangladesh for their hard work on this project. We gratefully acknowledge funding from the ESRC-DID Growth Programme managed by the Economic and Social Research Council of the UK (grant ES/LO12146/1). Woodruff recognizes support from ERC Advanced Grant RMGPP. We would like to thank all participants of the Institutional and Organisational Economics Academy, the Empirical Management Conference, the Jobs and Development Conference on Better Jobs for Development, the Oxford Firms, Labour & Organisations Research Group, and the CSAE Workshop for insightful comments and feedback. Ethical approval was obtained on the 30th of May 2014 from the University of Warwick Research Ethics Committee with reference 83/13-14.

Appendices of Chapter 2

2.A Description of the intervention

2.A.1 Training overview and timeline

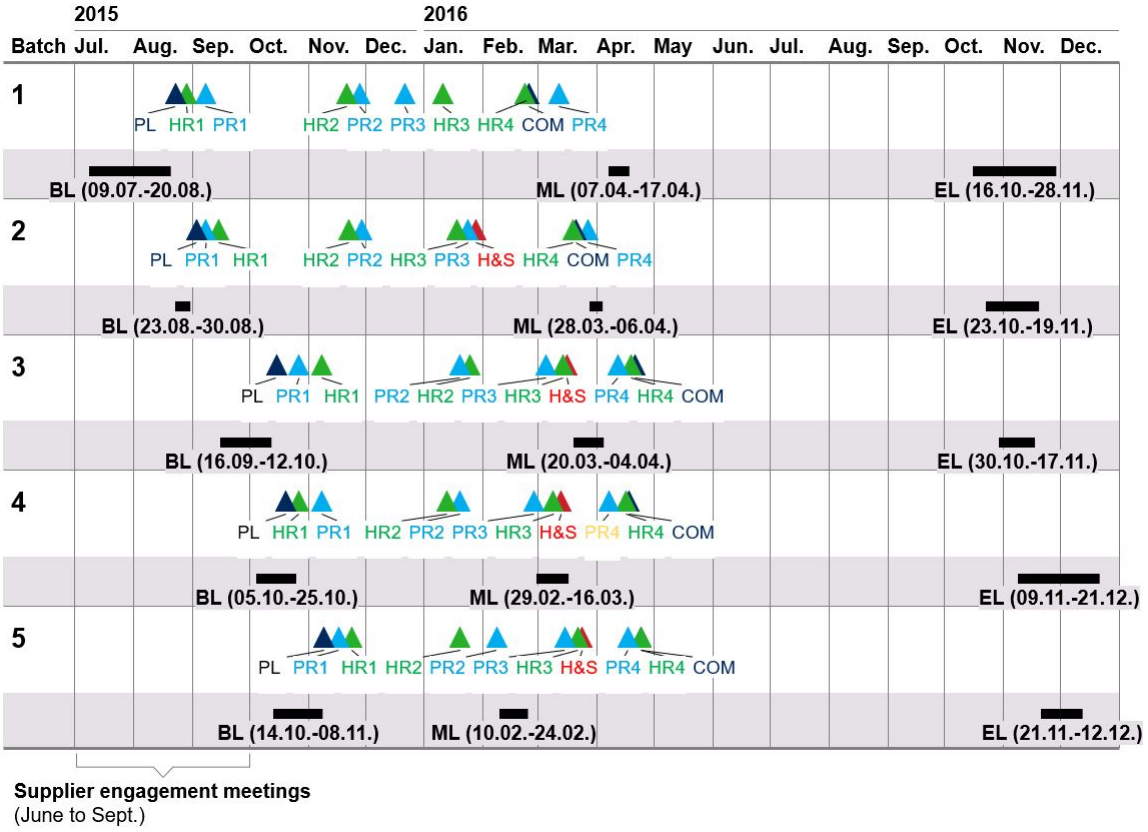


Figure 2.A.1: Timeline

The training was based upon a programme developed by Deutsche Gesellschaft für International Zusammenarbeit ("GIZ"). It is called the „Benefits for Business and Workers initiative", a management consulting service which offers training and practical expert advice to key factory managers on productivity and human resource management practices. The initiative was jointly carried out by Impactt Ltd, a UK-based consultancy, and Rajesh Bheda Consulting, an Indian-based consultancy focusing on enhancing the competitiveness in the fashion industry. The programme campaigns with the statement that "BBW helps transform factories into good businesses providing great jobs for their workers by developing a skilled, well-paid, safe and loyal workforce". On the one hand, businesses should have benefits from the programme, e.g. reduced absenteeism and worker turnover, more motivated workers and managers, increased productivity, savings through quality improvements and reduced wastage. On the other hand, workers should have benefits from the programme, e.g. increased job satisfaction, ability to raise and resolve issues, more skills, access to promotion, higher income and

lower working hours. Hence, it does not only consist of technical interventions to directly improve productivity, but also of HR interventions to indirectly increase productivity, e.g. via lower absence rates and lower turnover resulting of enhanced well-being of workers. The overall aim of the programme is to develop solid and robust systems to make sustained productivity and quality improvements, while improving worker's job quality.

The programme provided 24 days of group classroom-based training plus 4 follow-up visits and a roll-out session extended over a period of 6-8 months. The follow-up visits were scheduled regularly between training sessions to review progress and clarify content. The 1-day roll-out session served to roll out what they have learned to three other lines within the factory. At the end of the programme, factories could continue to roll out the most effective strategies on further lines.

Figure 2.A.1 shows a timeline with survey and training dates. Before each training start date, Baseline surveys were conducted (abbreviated as "BL"). Around the period of the third and fourth modules, midlines surveys were conducted (abbreviated as "ML"). Endline surveys were carried out approximately six months after the last training date (abbreviated as "EL"). The order of the training modules slightly varied from batch to batch, e.g. batch one and four were first taught human resource module one (HR1), whereas batch two, three and five were first taught production module one (PR1). In total, the training consisted of a plenary sessions to kick off the program, four HR and four PR modules, a Health and Safety module, and a Communication module.

2.A.2 Detailed training content

In the plenary session (1 day), participants were supposed to get to know each other and understand the requirements of the program. The first productivity module (3 days) was aimed at strengthening industrial engineering concepts and basic quality tools. In particular, this module taught concepts about method study, skills matrix, capacity study, line balancing, up-skilling operators, computation of productivity and quality baseline performance, rework capturing and analysis, cut to ship losses and asked participants to develop future action plans. The first human resource module (3 days) was aimed at empowering the HR department and review HR basics. In particular, it taught concepts about one-step leave policies, absenteeism policies, attendance bonus, buddy systems, exit interviews and exit registers, and instructions to track key HR metrics. The second productivity module (3 days) aimed at strengthening problem solving abilities and initiating Zero Defect Operators. Participants learned about methods to capture loss time, 5 Why approach, incl. cause and effect diagrams, internal customer-supplier orientation, and zero defect initiatives. The second human resource module (2 days) aimed at empowering supervisors to take on more HR responsibilities and improving teamwork. Specific training contents were job descriptions for workers and supervisors, team work, positive attitude towards work and root cause analysis, promotion and appraisal system for workers and supervisors, ways to engage middle managers and supervisor skills and training development. The Fire

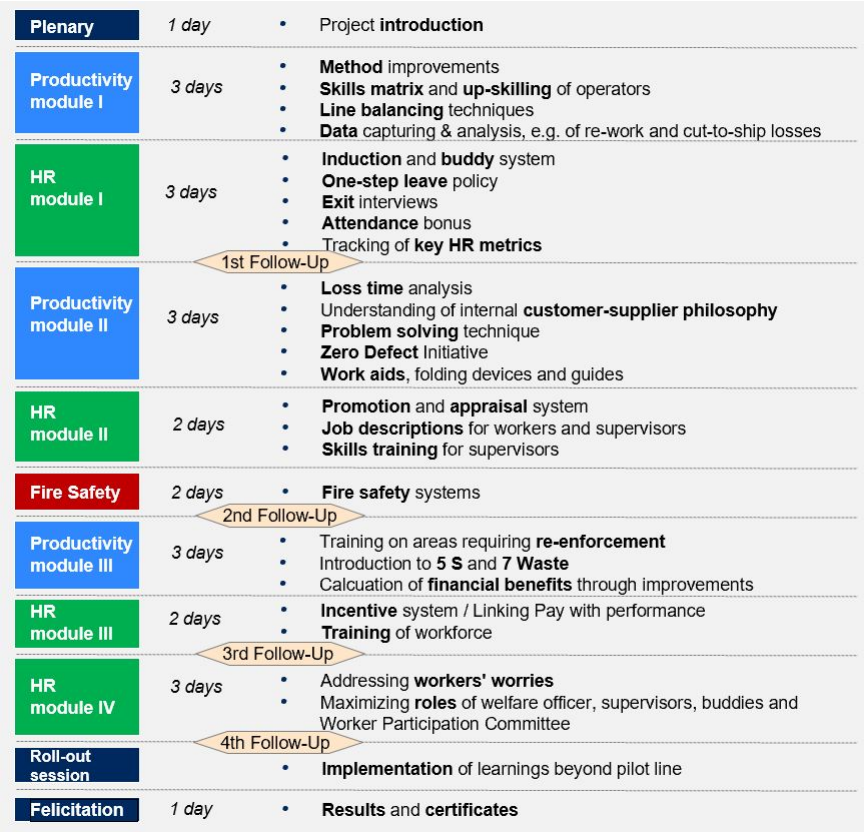


Figure 2.A.2: Training modules

and Safety module (2 days) was aimed at improving problem solving skills and how to make better decisions. It taught about current fire safety systems, barriers to effective systems, what a great safety system looks like, tools and techniques for strengthening current systems, what to consider in building safety and how to make better decisions. The third productivity module (3 days) was aimed at teaching computation of financial benefits through improvements and need based training formats. In particular, it introduced methods to calculate improvements and financial benefits, gave training on areas requiring reinforcement, and taught the 5S workplace organisation method (Sort, Set in Order, Shine, Standardize and Sustain), one of Toyota's lean manufacturing concept. The third human resource module (2 days) was aimed at concepts to get the best out of workers and provide them with the right skills, i.e. how to train workers, to develop a training plan, to decide upon promotions and link pay to performance, to develop production incentives and financial calculations. Participants also had to come up with a draft for improved production incentive schemes. The fourth human resource module (2 days) was aimed at providing managers with the right piece of mind, i.e. how to address workers' worries and how to maximize the role of welfare officer, supervisors, buddies and worker participation committee. It also taught how to train workers and representatives of the worker participation committee. Participants were required to share progress and challenges at the beginning and to develop fu-

ture action plans at the end of each module. Follow-ups were conducted four times (1 day each, so 4 days in total) to ensure a smooth implementation and to clarify any doubts. During the roll-out session (1 day), participants had to agree about the best ways to roll out the learnings of the pilot line to the rest of the factory. The felicitation session (1 day) was aimed at reviewing the progress, celebrating success, distributing certificates and agreeing upon forward action plans. Please see a summary of the modules in figure 2.A.2.

2.A.3 Evaluation of practices

Each practice taught during the training programme is evaluated by industry experts to assess the costs associated with implementation and operationalisation, in particular set-up costs, operating/maintenance costs and speed and precision of feedback/effects as shown in figure 2.A.3. Criteria to evaluate these costs are as follows:

- **Set-up costs**
  - Number of departments involved
  - Degree of standardization/formalization from low (not standardized), medium (somewhat standardized) to high (very standardized)
  - Resources required for set-up in terms of time from low (0-2 days), medium (2-10 days) to high (>10 days)
- **Operating/maintenance costs**
  - Number of users of the practice
  - Core skillset (soft or hard skill)
  - Degree of self-sustainability/frequency of required effort from low (continuous effort), medium (from time to time) to high (one-time effort)
- **Speed and precision of feedback**
  - Speed of feedback from immediate (within one week), short-term (within one month) to long-term (more than one month)
  - Measurability of effects from easy, medium to difficult

Points are given for each criteria, which are summed up to assess the rank across all practices.

				Set-up costs			Operating/maintenance costs			Speed and precision of feedback/effects	
Category	**Practice**	**Description**	**Rank**	**Number of depart-ments involved**	**Degree of standar-dization**	**Resou rces for set-up**	**Number of users**	**Core skills et**	**Degree of self-sustaina-bility**	**Speed of feedback**	**Measu-rability of effects**
HR	Induction system	- Show important factory locations (e.g. fire exit, welfare office) to newly hired operators - Hold conversation with welfare officers on first day of joining to reduce any discomfort	1	HR	high	low	Welfare Officer, HR Manager, PR Manager, Line Chief, Supervisor	Soft skills	medium	short-term	easy
	Buddy system	- Assign newly hired operators to more experienced operators to discuss any work-related issues - Seat newly hired operators next to more experienced operators	2	PR, HR	high	low	PR Manager, Line Chief, Supervisor, Welfare officer, HR Manager,	Soft skills	medium	short-term	easy
	One-step leave policy	- Ask operators to directly request leave from supervisor to reduce bureaucracy and improve the balance between worker well-being and workplace demand	2	PR, HR	high	low	PR Manager, Line Chief, Supervisor, HR Manager, HR Executive	Soft skills	medium	short-term	easy
	Worker representatives	- Assign one operator as worker representative to communicate concerns to management from the workers' point of view	4	PR, HR	high	high	Supervisor, Line Chief, Production Manager, Quality Supervisor, Quality Manager, HR Manager, Welfare Officer	Soft skills	medium	Imme-diate	medium
	Promotion policy	- Introduce clear KPIs to identify eligible workers for promotion - Make promotion policy transparent to workers	9	PR, HR, IE	medium	high	HR Manager, HR Executive, PR Manager, Line Chief, Supervisor, Quality Manager, Quality Controller, Quality Supervisor, IE Manager, IE Executive	Hard skills	low	long-term	medium
	Job descriptions	- Develop job descriptions for subordinates - Periodically review job descriptions with manager to ensure an understanding of mutual expectations, and to assess the achievement of goals and objectives	7	PR, HR, IE, QL	high	high	PR Manager, HR Manager, Quality Manager, IE Manager	Hard skills	medium	long-term	difficult
	Exit interviews	- Conduct exit interviews with workers leaving the factory - Discuss outcomes of exit interviews and identify patterns - Make changes in practices as a result of exit interviews	1	HR	high	low	Welfare Officer, HR Manager, PR Manager, Line Chief, Supervisor	Soft skills	medium	short-term	easy
	Attendance bonus	- Introduce attendance bonus - Make attendance bonus policy transparent to workers	3	HR	high	high	Welfare Officer, HR Executive, HR Manager	Soft skills	high	long-term	easy
PR	Up-skilling	- Assess skills gaps and prepare training plan to fill the gap - Implement training plan and record progress to analyse performance and quality of the training	8	PR, IE, Training	medium	high	Supervisor, Line Chief, Production Manager, IE Manager, IE Executive, Training Manager	Hard skills	low	long-term	medium
	Skills tests	- Conduct skills test once a new worker joined and regularly with existing workers - Maintain skills matrices to identify training areas	5	IE	medium	low	Supervisor, Line Chief, Production Manager, Quality Supervisor, IE Manager, IE Executive	Hard skills	medium	long-term	difficult
	Quality practices	- Develop zero defect operators who ensure to maintain required quality levels - Maintain files of checked garments - Use visual aids for quality inspections	6	PR, QL	medium	high	Supervisor, Line Chief, Quality Supervisor, Production Manager, Quality Manager	Hard skills	low	long-term	medium
	Production-related practices and tools	- Use stopwatch - Change line layout and set targets - Solve bottlenecks	4	PR, IE	high	high	Supervisor, Line Chief, Production Manager, IE Manager, IE Executive	Hard skills	low	Imme-diate	easy
Commu-nication	Communication practices	- Communicate targets to workers - Discuss work-related matters with managers and workers, e.g. turnover, quality, efficiency, or up-skilling - Discuss issues that affect workers	9	PR, HR, IE, QL	medium	high	Supervisor, Line Chief, Production Manager, Quality Supervisor, Quality Controller, Quality Manager, IE executive, IE Manager	Soft skills	low	long-term	difficult

Figure 2.A.3: Evaluation of practices

2.A.4 Organisational structure in Bangladeshi garment factories

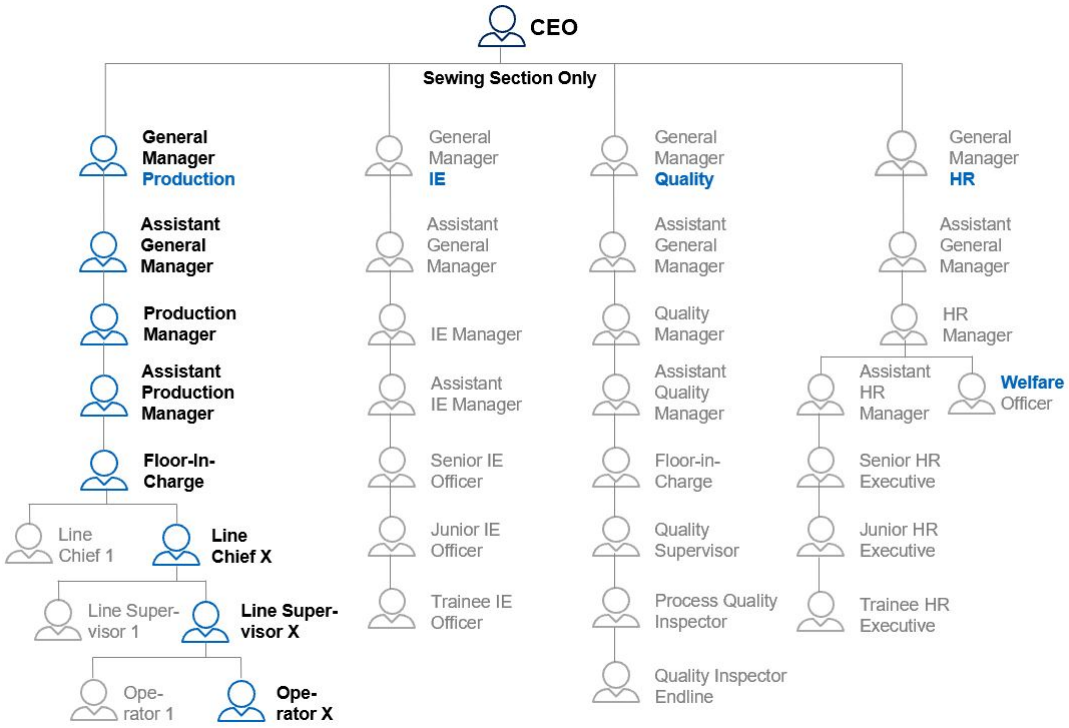


Figure 2.A.4: Chains of command

Notes: Figure 2.A.4 shows the management structure of a typical garment factory in Bangladesh. Note that not all factories have dedicated Industrial Engineering ("IE") departments.

Generally speaking factories specialise in the production of a particular family of garment. The sample of factories that participated in this project produced light knit and woven garments. These factories produce t-shirts, jeans, jackets and other similar products characterised by the type of fabric used. These factories do not produce knitted goods such as sweaters, nor goods utilising specialised techniques such as the metal work that is typical of lingerie producers. Whilst the method of organising the production of light knit and woven garments differs throughout the territories that produce those goods, it is fair to say that in Bangladesh the overwhelming model is that of the production line whereby cut cloth enters at the start of the line and moves sequentially with finished products exiting at the end of the line. Depending on the complexity of the garment the line will typically have between 15 and 70 sewing machine operators and five to 15 helpers, although there is considerable variation across factories in the ratio of helper to operators. The supervisory structure on the production lines tends to depend upon a variety of factors, the most pertinent of which is the number of machines. A line of 40 machines will typically have two supervisors and one line chief who may be in charge of one to three lines. The duties of line supervisors are various and include motivating workers, teaching workers new production techniques, identifying quality issues

and working with the IE department to solve problems that may cause bottlenecks on the line. Some factories have supervisors that work across multiple lines, and some do not have line chiefs; however, the majority of cases are as described above. Line supervisors and line chiefs are part of the production department and are directly supervised by the so-called Floor-In-Charge who is responsible for one production floor. In addition to the Production department, factories typically have Quality and HR departments. Quality supervisors/inspectors may be assigned to one production line although they generally have responsibility for more than one line. The Welfare Officer belongs to the HR department. Some factories also have dedicated Industrial Engineering ("IE") departments that manage the most technical aspects of production including setting targets and determining the layout of the line. Using IE techniques is certainly thought to be best practise, yet whilst the presence of IE departments is not uncommon they are far from ubiquitous. As the chains of command in figure 2.A.4 show, each department has a General Manager, an Assistant General Manager, a Departmental Manager, and an Assistant Departmental Manager. Due to steep hierarchical structures and lack of authority at lower managerial levels, instructions flowing downwards to workers or supervisors, and approvals/requests flowing upwards to general managers have to go through many layers of the organisation. Decisions tend to be made only at the highest managerial level.

2.A.5 Assignment into batches

The participating 25 factories were sorted into five “early” and “late” batches, with an early or late start date between August and November 2015. Batches one and two were classified as early batches and batches three to five were classified as late batches. The assignment was affected by two factors. Firstly, the composition of batch 1 with six factories had already been decided on a first-come-first-serve basis before the randomisation strategy was developed. Four out of those six were factories of the main sample, one was the pilot factory and one was a separate factory not included in the sample. Hence, only the remaining 18 factories (22 minus the four factories from the main sample which had already been assigned to Batch 1) could randomly be assigned into batch 2 to 5. Secondly, the country director of BBW ("Benefits for Business and Workers") programme in Bangladesh had a strong preference for grouping batches based on location for logistical reasons, reducing the flexibility to assign completely random. 12 factories out of 18, which were roughly in the same location, were randomly assigned into an early and into a late batch. The other 6 factories were automatically placed in the late batch. In September 2015, four new factories expressed interest in joining the programme. They were assigned into a new batch 5. One final factory, which decided to join the programme end of October, had to be assigned to batch 5 as well, as the start date of previous batches was already due. In September 2015, one factory dropped out due to closure of the factory and in October 2015, one additional factory dropped out due to internal problems, reducing the total sample to 25.

To summarise (see figure 2.A.5), only 12 factories located closely to each other

Batch	Nb. of factories	Assignment	Training start date
Batch 1	Four (+ 2 non-sample)	'First come, first served' basis	August 23, 2015
Batch 2	Six	Random (similar location)	September 3, 2015
Batch 3	Five	Random (similar location)	October 15, 2015
Batch 4	Five	By default (diff. location)	October 20, 2015
Batch 5	Five	By default (late join date)	November 9, 2015

Figure 2.A.5: Assignment into batches

were randomly assigned into an early batch (batch two) or late batch (batch three). Batch one was decided before the randomisation strategy was developed on a 'first come, first serve' basis, batch four was reserved for remaining factories in different locations, and batch five for factories which joined later.

2.B Description of the data

2.B.1 Data overview

We use both detailed factory data and survey data for the analysis in this paper. Factory data is collected from the factories at least three months before the first training and continuing for a period of at least six months after the last training. Importantly, we collect data on all lines in the participating factories, whether those lines were nominated for training or not. Thus, we have a very clean randomised control group against which to measure the short-run effects of training (nominated/surveyed lines only, i.e. treatment, adjacent and control lines), and can also obtain credible estimates of effects using a set of comparison lines that were not selected by the factories to receive training (non-surveyed lines). Factory data consists of two types of data: Production data and salary data. Production data is collected at the daily level and includes line-level productivity-related information such as styles produced, output, allocated and actual manpower in numbers, manpower in available minutes, order quantity, running days, or standard minute value, and quality-related information such as number of spots, rejects, and alterations, or number of pieces quality checked. The so-called standard minute value (smv) is a style-specific value, calculated prior to the start of production of the style. It is the sum of the time, in seconds, it takes to perform each sewing operation to assemble one piece of the style. Thus, it provides a measure of the required labour input per piece under ideal production conditions. We collapse production data at the weekly level. Salary data is collected at the monthly level and includes worker-level information such as grades, attendance, absenteeism, days paid, salary deductions or allowances. We drop the month of the first training in our analysis to ensure a clean difference between pre- and post-training months.

The surveys were conducted at three points of time: Just before the first training round, 2-3 months after the first training round, and six months after the last training round. Survey respondents were operators, line supervisors/line chiefs and managers.

2.B.2 Main outcomes of interest using factory data

Index	Variable	Definition
HR	Exit rates	Nb. exits within 2 months / Nb. workers joining per month
	Turnover rates	Nb. exits after 2 months / Nb. workers
	Absenteeism	Nb. workers absent / Nb. workers absent + Nb. workers present
	Promotion rates	Nb. promotions / Nb. workers
PR	Efficiency	Output minutes (smv*daily output) / Available minutes
	Alteration rates	Nb. alterations / Nb. pieces checked
	Reject rates	Nb. rejects / Nb. pieces checked

Figure 2.B.6: Definition of outcome variables

The factory data allows us to analyse two indices, an index related to human resource outcomes and an index related to production. For the HR index, we use monthly salary data to construct exit rates, turnover rates, absenteeism and promotion rates. For the PR index, we use weekly production data to construct efficiency, alteration rates and reject rates. Both indices are generated as the sum of each index component divided by the total number of index components based on Kling, Liebman & Katz (2007). The exact definitions of these variables are shown in figure 2.B.6.

2.B.3 Main outcomes of interest using survey data

The survey data allows us to analyse four indices, an index related to human resource practices, production practices, worker wellbeing, and communication practices. All four indices are constructed using Anderson's method of variance-covariance weighted indices (2008), and are available for each type of survey respondents: Line operators, line supervisors/line chiefs, and managers. The type of questions asked differed per respondent type. A detailed overview of the sub-indices included in each index, as well as a description of all variables is given in figure 2.B.7 for line operators, 2.B.8 for line supervisors/line chiefs, and 2.B.9 for managers.

Index	Sub-Index	Variable	Description
HR	Induction system	Induction	Worker was shown none, several or all of various important factory locations (Production Manager's office, Welfare office, Fire exit)
		Welfare officer	Welfare officer had conversation with worker during first week
	Buddy system	Paired	Operator was sitting next to experienced worker during first months at factory
		1st talk to Buddy	Buddy is first person to talk to if worker has work-related grievance
		Mentor useful	Buddy is most helpful during first weeks at the factory
		Have mentor	Buddy exists on the line
		Be mentor	Worker has even been assigned as mentor
		Questions	Frequency of being asked questions by new operators
		Teaching ability	Assessment of teaching and mentoring ability
	One-step leave policy	One-step	Only one person approached to request leave
		Time away	Time spend away to request leave
	Worker representative	Work rep known	Operator knows who worker representative is
		Satisfied	Frequency of adequate communication of concerns to management
	Promotion policy	Promotion requirements	Operator knows requirements for promotion
		Promotion beliefs	Worker believes effort or effort plus supervision is most important in determining the likelihood and timing of a promotion
	Attendance bonus	Amount of bonus	Knowledge of amount of attendance bonus at the factory
		Eligibility	Knowledge about eligibility of bonus (if 5 min. late for one day of the month or if 1 day announced leave taken)
PR	Up-skilling	Improve	Receive help in improving sewing machine skills
		Time and who	Duration, frequency and people involved in helping workers to improve sewing machine skills
		New skill	Belief regarding effect of mastering new skill in skill test
		Increased efficiency time	Time spent by supervisor or line chief instructing worker on how to perform sewing processes more efficiently last week
	Skills test	Test frequency	Frequency of formal skills test by superior
		Tests pattern	Is test done regularly or upon request
		Last test	Last time skills were tested
		Test duration	Duration of test
		Next test	Worker was told when next test will be
		Test feedback	Worker was given feedback on test
Wellbeing	Feelings of surveyed group	Conflict with MG	No perceived conflict between operators and managers
		Self-Assess.	Assessment of own feelings towards job
		Others	Assessment of feelings of others towards job
	Involvement of workers	Change	Own responsibility for changes on the line
		Sugg	Operator involvement in idea generation on the line
Commu-nication	Communication practices	Initiate	Variety of topics of initiated meetings with immediate superior (production capabilities, skills and/or work grievances)
		Approach	Frequency of approaching supervisors to tell them about issues that affect workers
		Discuss	Frequency of talking to other operators about how to make things better at the factory
	Knowledge about LS/LC	1st position	First position in the garment industry
		Married	Indicator for marriage
		Length	Length of marriage
		#Children	Number of children
		#Boys	Number of boys
		#Girls	Number of girls
		District	Home district
		Division	Home division
		Years	Tenure in the factory in years
		Months	Tenure in the factory in months

Figure 2.B.7: Components of line operator indices

Notes: Figure 2.B.7 shows all sub-indices of the line operators' HR, PR, Wellbeing and Communication Index, including a detailed description of the components of the sub-indices.

Index	Sub-Index	Variable	Description
HR	One-step leave	Leave request last	Last time an operator requested to leave
	Job descriptions	Description	Ever seen copy of job description
		Description review	Last time job description was reviewed with managers
	Attendance bonus	Amount of bonus	Knowledge of amount of attendance bonus at the factory
		Eligibility	Knowledge about eligibility of bonus (if 5 min. late for one day of the month or if 1 day announced leave taken)
PR	Up-skilling	Time spent last week	Time spent improving operator skills last week
		Time spent normally	Time spent improving operator skills normally
		Skills to improve	Decision about which skills to improve
	Skills test	Skills matrix used	Skills matrix used to rebalance a line/resolve a bottleneck
		Availability of skills matrix	Skills matrix for operators available
		Involvement in update	In update of skills matrix involved
		Frequency of update	Frequency of updates of skills matrix
		Frequency of tests	Frequency of skills tests
		Regularity of tests	Tests scheduled regularly or upon request
	Quality practices	Changes quality checks	Change to quality checks in last 6 months
		What changes	What were these changes
		Inspection	Line uses endline inspection format
		File checks	Line maintains a file for tracking quantity of garments checked
		Flags	Flags used as method/aid in quality inspection
	Production-related practices	Efficiency level	Efficiency level on last day worked
		Stopwatch	Stopwatch used in the last month to time how long operator requires to complete task
		Job assignments	People involved in assigning operators to specific tasks
		Line layout	Collaborate with IE department or responsible people to design line layout
		Targets	In last style change, involved by IE department or people responsible in setting targets
		Efficiency of processes	Ever identified a process on current style that could be performed better with fewer motions
		Movement reduction	Ever attempted to reduce the movement required for operator to complete specific process
		Bottleneck solution	Select most probable solution offered to solve bottleneck on LSLC's line
		Efficiency training	Received training on how to improve efficiency in last 2 months
		Location efficiency training	Location of efficiency training
	Production-related tools	Bring to style meeting	What does LSLC bring to meeting on upcoming style
		Bring to performance meeting	What does LSLC bring to meeting on operator performance
		Bring to target meeting	What does LSLC bring to meeting on line's daily target
		Bring to individual targets meeting	What does LSLC bring to meeting on upcoming style
Wellbeing	Feelings of surveyed group	Collaboration	Agree/disagree that there will always be conflict between managers and workers
		Self	Attitude towards own work
		Peer	Attitude towards/belief about other supervisors' work
	Involvement of workers	Suggestions	Expectation in the factory that workers should make suggestions for improving the production process
Commu-nication	Communication practices	Initiate	In past month, initiated meeting with immediate superior to discuss none, one or several work related issues
		Targets	How are targets communicated to operators
		Meetings	Initiated meeting to discuss work related matters
		Training	Received communication training in last 2 months
		Location training	Location of communication training
	Knowledge about LS/LC	Section	Same section of the line
		Education	Education level
		Married	Indicator for marriage
		Length	Length of marriage
		#Children	Number of children
		#Boys	Number of boys
		#Girls	Number of girls
		Children's age	Age of youngest child
		Years	Tenure in the factory in years
		Months	Tenure in the factory in months
	Communication with industrial engineer	Styles	Frequency of meetings on upcoming style during the week before a new style
		Performance	Frequency of meetings on individual operator performance during the week before a new style
		Targets	How are targets communicated to operators
		Operator targets	Frequency of meetings on targets for individual operators during the week before a new style
		Organization	Meetings regularly scheduled or on needs-basis

Figure 2.B.8: Components of line supervisor/line chief indices

Notes: Figure 2.B.8 shows all sub-indices of the line supervisor/line chiefs' HR, PR, Wellbeing and Communication Index, including a detailed description of the components of the sub-indices.

Index	Sub-Index	Variable	Description
HR	Buddy system	Buddy in line	Indicator for whether there is buddy on pilot line
		Number of lines with buddy	Number of lines with buddy
		Duration buddy system	Amount of time buddy system has been active
		Buddy issues	Indicator for whether buddies have approached the manager about worker issues
		Buddy issues frequency	Frequency of approaches by buddies to managers about worker issues
	One-step leave policy	Involved	Indicator for whether respondent is from HR and thus involved in approving leave
		Leave go-to person	Persons approached by operators in order to request leave
		Direct request	Indicator for whether operators request leave to respondent directly
		Indirect request	Indicator for supervisor requesting leave for operators
	Promotion policy	Promotion policy	Indicator for whether there is written promotion policy in factory
		Promotion awareness	Percentage of operators aware of promotion policy
		Promotion resources	Number of resources used to identify workers that should be promoted
		Percent promotions	Percentage of operators that get promotions in a year in the factory
		Promotion trend	Increase in number of promotions
	Exit interviews	Last exit	Indicator for time passed longer than one week since a worker left
		Numbers of exits	Indicator for less than five people left on that day
		Exit interview	Indicator for conducted exit interview with that person
		Percentage	Percentage of all leaving workers that had exit interview
		Duration	Duration of typical exit interview in minutes
		Format	Indicator for fixed format of exit interview
		Discussants	Number of people that discuss/have access to results
		Change	Indicator for change in practice since last exit interview
		Any changes	Indicator for any change in practices as a result of exit interviews
		Timing	Timing of last meeting to discuss exit interviews
		Follup up	Frequency of following up to see if change were actually made
	Attendance bonus	Awareness	Share of operators aware of attendance bonus
PR	Up-skilling	Involved in improvements	Indicator for whether manager is involved in improving operator skills
		Upskilling method	Indicator for change in method to improve worker skills
		Changes quality checks	Number of positive changes to such methods
		Upskill time	Time spent upskilling per week
		Upskill time 6 months ago	Time spent upskilling per week 6 months ago
		Upskilling location	Number of locations (e.g. off the line and on a running line) where upskilling takes place
	Skills test	Skills matrix	Indicator for whether factory maintains skill matrices for operators
		Skills matrix update	Frequency of updates of skill matrix
		Skills matrix last use	Last usage of skill matrix
		Skill matrix used for	Number of decisions or processes for which skill matrices are used
	Quality practices	New system	Indicator for whether factory has implemented new quality system in past 6 months
		New system additions	Number of new practices/systems added
		0 defect operators	Number of 0 defect operators on pilot lines
		0 defect lines	Number of 0 defect lines in the factory
	Production-related practices	Target high	Degree of flexibility in the factory for target recalibration
		Last layout change	Time passed since last line layout change
		Solution bottleneck	Probable solution in case of bottleneck
		Resources bottleneck	Number of resource types that would be used in case of bottleneck
Wellbeing	Feelings of surveyed group	Self-Assess.	Assessment of own feelings towards job
		Assessment of peers	Assessment of feelings of IE/HR/PR/WF/QL manager
	Operators' feelings	Assessment of operators	Assessment of feelings of operators
Commu-nication	Communication practices	Index communication	Index including sub-indices about general meetings, turnover meetings, quality meetings, efficiency meetings, upskilling meetings (incl. frequency, participants, duration and regularity)

Figure 2.B.9: Components of manager indices

Notes: Figure 2.B.9 shows all sub-indices of the managers' HR, PR, Wellbeing and Communication Index, including a detailed description of the components of the sub-indices.

2.B.4 Attendance

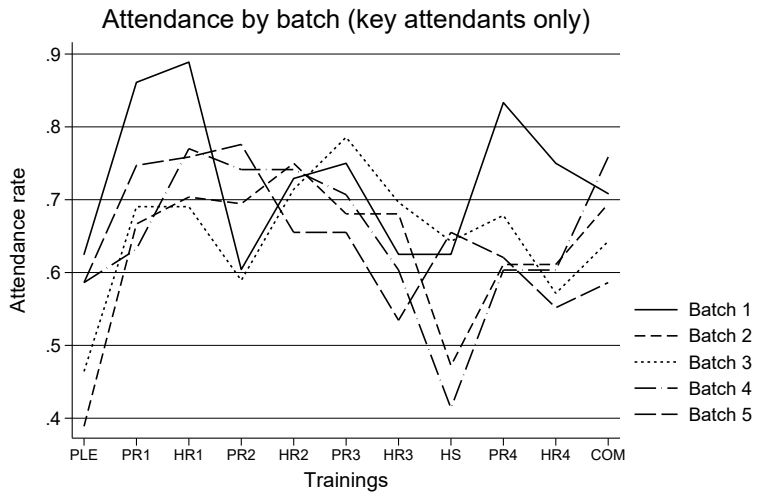


Figure 2.B.10: Attendance of key attendants by batch

Notes: Figure 2.B.10 shows the attendance rate of key attendants by batch. Factories are grouped into five batches as outline in appendix Section 2.A.5. Key attendants are defined by the training provider. They are middle and/or lower managers at which the respective training modules are targeted.

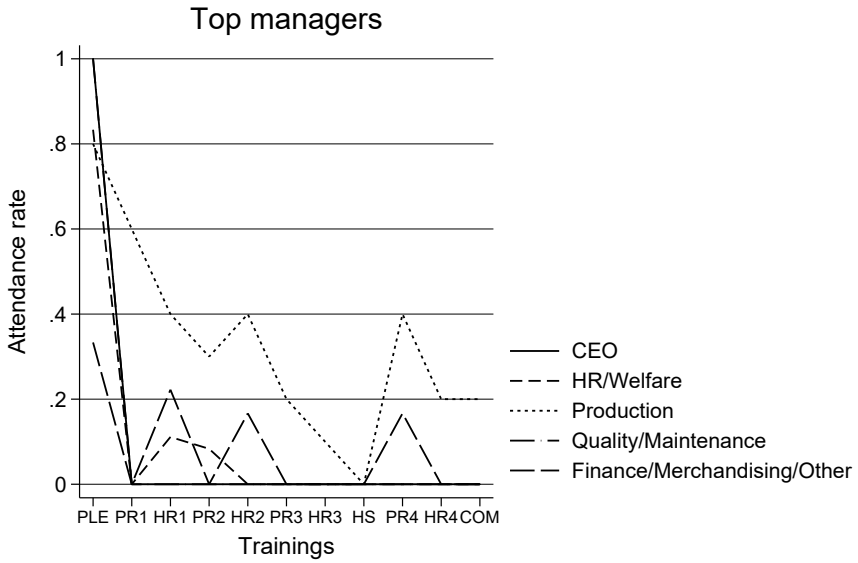


Figure 2.B.11: Attendance of top-level managers by department

Notes: Figure 2.B.11 shows the attendance rate of high-level managers, i.e. CEO (General Managers and Assistant General Managers) and the heads of departments.

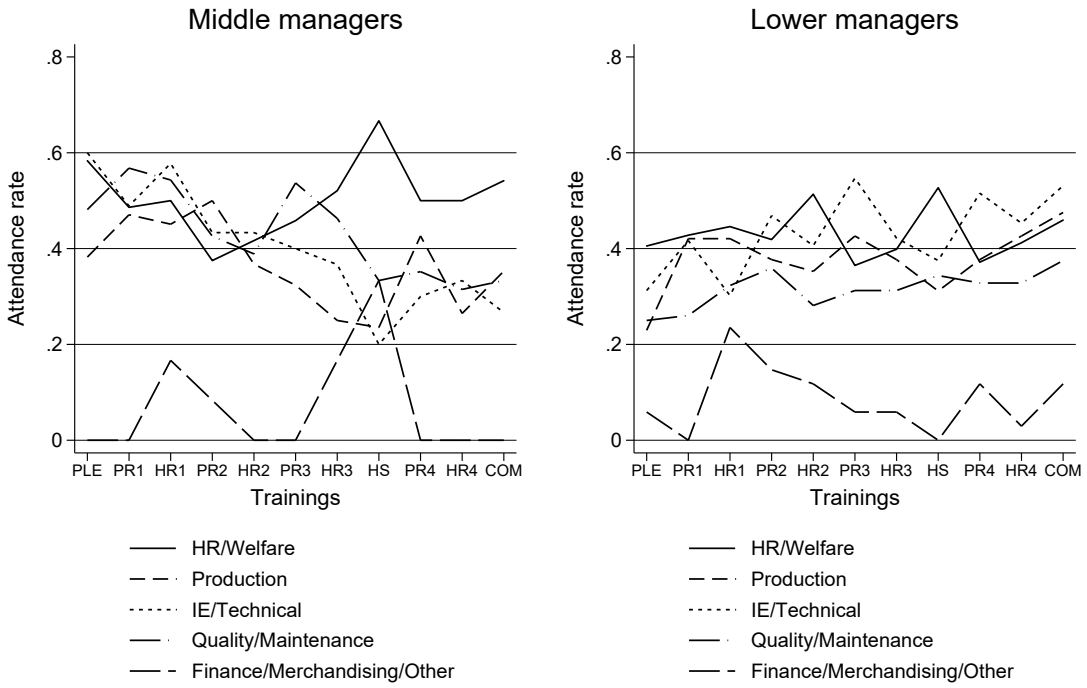


Figure 2.B.12: Attendance of middle- and low-level managers by department

Notes: Figure 2.B.12 shows the attendance rate of middle managers in the left panel and low-level managers in the right panel. Middle managers include Deputy Managers, Assistant Managers or Managers, whereas lower managers include Officers, (Senior) Executives, Supervisors or Floor-In-Charges.

2.C Empirical specifications

2.C.1 Factory types

We split the total number of 25 factories into three different types of factories depending on the location of control lines. 14 factories are assigned to type A, which has control lines only on different floors than treatment lines. 3 factories are assigned to type B, which has control lines on both the same and a different floor than treatment lines. The remaining 8 factories are assigned to type C, which has control lines only on the same floor as treatment lines. The reason to classify factories as such is that it allows us to run regressions for different types of factories and to test our hypothesis of larger spillover effects to control lines on the same floor as treatment lines. We call production floors with one randomly selected treatment line as "treatment floor", and floors without treatment lines, but with at least one nominated non-selected line as "control floor".

2.C.2 Specifications for survey data

		All C-lines		C-lines on same floor	C-lines on diff floor
		Without adjacent lines	With adjacent lines		
Floor	Lines	(1)	(2)	(3)	(4)
Treatment	Treatment	T T T	T	T	T
	Control	C C	C	C	
	Adjacent		A	A	A
Control	Control	C C	C		C
Factory types		A + B + C	ABC	BC	AB
#Factories		14 + 3 + 8	25	11	17

Figure 2.C.13: Illustration of empirical specifications: Worker outcomes

Figure 2.C.13 illustrates which types of factories and lines are included in each empirical specification using survey data. The specification illustrated in the first column excludes adjacent lines, whereas the specifications illustrated in columns two to four include adjacent lines. Data from all three factory types is used in columns one and two, whereas only "BC" factories (with control lines on the same floor as treatment lines) and "AB" factories (with control lines on different floors than treatment lines) are included in columns three and four, respectively. Columns one and two correspond to regressions shown in Table 2.5 and 2.6. The third column refers to regressions indicated as "Same" in Table 2.7, as well as Tables 2.D.1 and 2.D.2. The fourth column refers to regressions indicated as "Diff" in Table 2.7, as well as Tables 2.D.3 and 2.D.4.

The variables used for the empirical specification, $TxML$, $TxEL$, $AxML$, and $AxEL$, to analyse data from line operators and line supervisors/line chiefs are il-

lustrated in figure 2.C.14. T is an indicator equal to one for treatment lines, A is an indicator equal to one for adjacent lines, and ML and EL are indicators equal to one for midline resp. endline data. As managers are responsible for several production floors with treatment, adjacent and control lines, the analysis of manager data does not allow us to differentiate between responses by line. However, we can analyse the manager data over time, and see whether responses differ between midline and endline.

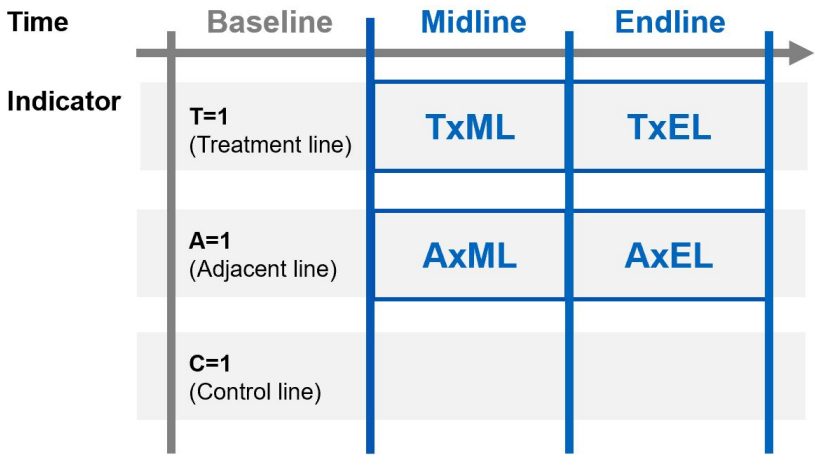


Figure 2.C.14: Illustration of independent variables: Worker outcomes

2.C.3 Specifications for factory data

Figure 2.C.13 illustrates which types of factories and lines are included in each empirical specification using factory data. Whereas survey data is only available for surveyed lines, factory data is collected for surveyed (treatment, adjacent and control lines) and non-surveyed lines on both treatment and control floors. We argue that non-surveyed lines on treatment floors are similar in expectation to non-surveyed lines on control floors, allowing us to include factory data from both surveyed and non-surveyed lines in our analysis. The first three specifications (1 to 6) compare treatment to control lines. The first column "All C-lines (ABC)" (1 and 2) indicates that all control lines in factories of all types are included. The second column "All C-lines (AB)" (3 and 4) indicates that only factories of type A and B with at least one control line on a different floor than the treatment line are included. We can then compare these results to column "Diff C-lines" (5 and 6) with control lines on different floors only. Specifications 7 to 12 include adjacent lines to better analyse spillover effects. The fourth column (7 and 8) indicates the comparison of treated and adjacent lines to control lines in all factories, whereas the fifth column (9 and 10) excludes factories with only control lines on the same floor as the treatment line, and only keeps control lines on different floors. The sixth column (11 and 12) shows the comparison of treated, adjacent and all other lines on the treatment floor in AB factories against the control floors. The last two

columns (13 to 16) show a floor comparison, with a comparison of the treatment floor to either just the control floor (13 and 14) or to all floors in the whole factory (15 and 16). Summary Table 2.9 shows the corresponding regression results for each of the specifications as described in figure 2.C.15 (the specification numbers in figure 2.C.15 from one to sixteen match with the column numbers in the summary regression Table 2.9). Regressions results of individual survey components for each specification are shown in Tables 2.E.16 to 2.E.22.

Floor	Lines	All C-lines *ABC Type* (1)(2)	All C-lines *AB Type* (3)(4)	Diff C-lines *AB Type* (5)(6)	Adj-lines *ABC Type* (7)(8)	Adj-lines *AB Type* (9)(10)	Adj+Other *AB Type* (11)(12)	C-floor *AB Type* (13)(14)	All floors *AB Type* (15)(16)
Treat-ment	Treatment	T T T	T	T	T	T	T	T	T
	Control	C C	C		C		Other	T	T
	Adjacent				Adjacent	Adjacent	Adjacent	T	T
	Non-survey						Other	T	T
Control	Control	C C	C	C	C	C	C	C	C
	Non-survey						C	C	C
Remaining	Non-survey								C
Factory types		A + B + C	AB	AB	ABC	AB	AB	AB	AB
#Factories		14 + 3 + 8	17	17	25	17	17	17	17

Figure 2.C.15: Illustration of empirical specifications: Firm outcomes

The variables used for the empirical specification, $PostX$ and $TxPostX$ (for X=1, 2, 3, 4), are illustrated in figure 2.C.16. $PostX$ is an indicator variable for the time period between the Xth and X+1th training. $Post4$ is an indicator variable for three months after the last training date. All $PostX$ variables are interacted with the treatment variable. We use HR training dates for HR indices and PR trainings dates for PR indices.

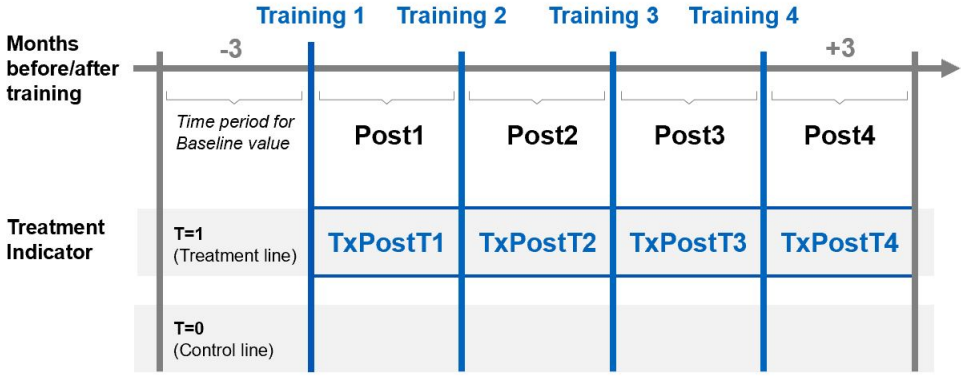


Figure 2.C.16: Illustration of independent variables: Firm outcomes

2.D Supplementary figures and tables: Worker outcomes

TABLE 2.D.1. HR AND PR INDEX: CONTROL LINES ON SAME FLOOR

	HR Index					PR Index				
	(1) LO	(2) LO	(3) LSLC	(4) LSLC	(5) MG	(6) LO	(7) LO	(8) LSLC	(9) LSLC	(10) MG
Endline	0.10	0.018	-0.37	-0.56		0.29	0.43	-0.96*	-0.79**	
	(0.44)	(0.41)	(0.95)	(0.80)		(0.50)	(0.38)	(0.49)	(0.39)	
MidlineXTreated	0.23	0.094	0.19	-0.17		-0.026	-0.13	0.49	0.45	
	(0.44)	(0.40)	(0.65)	(0.37)		(0.32)	(0.35)	(0.73)	(0.44)	
EndlineXTreated	-0.089	-0.16	0.29	0.17		-0.27	-0.57*	1.23*	0.96*	
	(0.45)	(0.44)	(0.69)	(0.68)		(0.40)	(0.30)	(0.64)	(0.50)	
MidlineXAdjacent		-0.10		-0.71**			-0.63*		-0.44	
		(0.41)		(0.33)			(0.34)		(0.40)	
EndlineXAdjacent		-0.36		0.074			-0.81**		0.33	
		(0.42)		(0.65)			(0.34)		(0.35)	
Treated					0.24**					-0.32**
					(0.11)					(0.12)
Constant	0.37	0.57*	0.23	0.48*	-0.11**	0.22	0.27	0.18	0.090	0.15**
	(0.34)	(0.32)	(0.48)	(0.25)	(0.053)	(0.22)	(0.28)	(0.56)	(0.33)	(0.056)
Observations	76	159	76	165	247	76	159	76	165	247
Line clusters	38	81	39	88	25	38	81	39	88	25
p: MLxT=MLxA		0.51		0.075			0.073		0.0093	
p: ELxT=ELxA		0.49		0.72			0.43		0.071	

Notes: Table 2.D.1 shows the main treatment effects on the HR index and PR index excluding control lines on a different floor than the treatment lines. All indices are based on Anderson (2008). Index components are different for Line operators (LO), Line supervisors/line chiefs (LSLC) and managers (MG) (see Section 2.B.3 in the appendix). Midline and Endline are indicator variables equal to one if responses are from midline and endline surveys, respectively. Adjacent lines are lines next to treatment lines. MidlineXTreated and EndlineXTreated are indicator variables equal to one if both the treatment indicator and the respective survey indicator are equal to one. MidlineXAdjacent and EndlineXAdjacent are indicator variables equal to one if the adjacent line indicator and the respective survey indicator are equal to one. Regressions include factory fixed effects and the error term is clustered at the line level. p-values are shown at the bottom of the table for testing the equality of coefficients on treatment and adjacent lines at midline and endline (MLxT=MLxA; ELxT=ELxA). Standard errors are shown in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE 2.D.2. WELLBEING AND COMMUNICATION INDEX: CONTROL LINES ON SAME FLOOR

	Wellbeing Index					**Communication Index**				
	(1) LO	(2) LO	(3) LSLC	(4) LSLC	(5) MG	(6) LO	(7) LO	(8) LSLC	(9) LSLC	(10) MG
Endline	0.55	0.42	0.67	0.82		0.99**	1.00**	-1.02	-0.74	
	(0.45)	(0.39)	(0.88)	(0.73)		(0.48)	(0.44)	(0.73)	(0.57)	
MidlineXTreated	0.032	-0.31	0.19	0.15		1.35***	1.22***	-0.46	-0.33	
	(0.32)	(0.33)	(0.63)	(0.63)		(0.38)	(0.34)	(0.52)	(0.34)	
EndlineXTreated	0.70**	0.53*	0.038	-0.14		0.95***	0.77**	0.88	0.65	
	(0.33)	(0.32)	(0.52)	(0.47)		(0.35)	(0.38)	(0.69)	(0.59)	
MidlineXAdjacent		-0.022		-0.077			1.47***		-0.84**	
		(0.33)		(0.61)			(0.38)		(0.35)	
EndlineXAdjacent		0.39		-0.31			0.30		0.58	
		(0.33)		(0.50)			(0.40)		(0.50)	
Treated					-0.039					0.14
					(0.16)					(0.16)
Constant	-0.53**	-0.27	-0.22	-0.25	0.018	-0.49*	-0.38	0.67	0.57**	-0.065
	(0.21)	(0.25)	(0.58)	(0.55)	(0.073)	(0.29)	(0.29)	(0.40)	(0.29)	(0.072)
Observations	76	159	76	165	247	78	159	76	165	247
Line clusters	38	81	39	88	25	40	81	39	88	25
p: MLxT=MLxA		0.33		0.41			0.42		0.016	
p: ELxT=ELxA		0.64		0.56			0.13		0.85	

Notes: Table 2.D.2 shows the main treatment effects on the Wellbeing and communication index. Control lines are only included if they are located on the same floor as treatment lines. All indices are based on Anderson (2008). Index components are different for Line operators (LO), Line supervisors/line chiefs (LSLC) and managers (MG) (see Section 2.B.3 in the appendix). Regressions include factory fixed effects and the error term is clustered at the line level. p-values are shown for testing the equality of coefficients on treatment and adjacent lines at midline and endline (MLxT=MLxA; ELxT=ELxA). Standard errors are shown in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE 2.D.3. HR AND PR INDEX: CONTROL LINES ON DIFFERENT FLOOR

	HR Index					**PR Index**				
	(1) LO	(2) LO	(3) LSLC	(4) LSLC	(5) MG	(6) LO	(7) LO	(8) LSLC	(9) LSLC	(10) MG
Endline	-0.14	-0.31	0.085	0.025		-0.18	-0.14	0.23	0.12	
	(0.41)	(0.35)	(0.25)	(0.23)		(0.35)	(0.30)	(0.30)	(0.28)	
MidlineXTreated	0.56	0.70**	0.24	0.14		0.29	0.35	0.41	0.37	
	(0.36)	(0.32)	(0.24)	(0.25)		(0.35)	(0.33)	(0.26)	(0.29)	
EndlineXTreated	0.47	0.78**	-0.14	-0.15		0.51	0.48	-0.12	-0.082	
	(0.35)	(0.35)	(0.32)	(0.30)		(0.38)	(0.34)	(0.37)	(0.39)	
MidlineXAdjacent		0.50*		-0.42*			-0.16		-0.52*	
		(0.30)		(0.24)			(0.33)		(0.29)	
EndlineXAdjacent		0.58*		-0.24			0.25		-0.65**	
		(0.30)		(0.27)			(0.36)		(0.29)	
Treated					0.24**					-0.32**
					(0.11)					(0.12)
Constant	0.047	0.036	0.025	0.12	-0.11**	-0.070	-0.17	0.054	0.098	0.15**
	(0.24)	(0.21)	(0.14)	(0.14)	(0.053)	(0.19)	(0.24)	(0.18)	(0.18)	(0.056)
Observations	104	187	108	193	247	104	187	108	193	247
Line clusters	52	95	54	100	25	52	95	54	100	25
p: MLxT=MLxA		0.52		0.046			0.088		0.0074	
p: ELxT=ELxA		0.49		0.71			0.46		0.075	

Notes: Table 2.D.3 shows the main treatment effects on the HR index and PR index excluding control lines on the same floor as treatment lines. All indices are based on Anderson (2008). Index components are different for Line operators (LO), Line supervisors/line chiefs (LSLC) and managers (MG) (see Section 2.B.3 in the appendix). Regressions include factory fixed effects and the error term is clustered at the line level. p-values are shown at the bottom of the table for testing the equality of coefficients on treatment and adjacent lines at midline and endline (MLxT=MLxA; ELxT=ELxA). Standard errors are shown in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE 2.D.4. WELLBEING AND COMMUNICATION INDEX: CONTROL LINES ON DIFFERENT FLOOR

	Wellbeing Index					**Communication Index**				
	(1) LO	(2) LO	(3) LSLC	(4) LSLC	(5) MG	(6) LO	(7) LO	(8) LSLC	(9) LSLC	(10) MG
Endline	0.55	0.42	0.67	0.82		0.99**	1.00**	-1.02	-0.74	
	(0.45)	(0.39)	(0.88)	(0.73)		(0.48)	(0.44)	(0.73)	(0.57)	
MidlineXTreated	0.032	-0.31	0.19	0.15		1.35***	1.22***	-0.46	-0.33	
	(0.32)	(0.33)	(0.63)	(0.63)		(0.38)	(0.34)	(0.52)	(0.34)	
EndlineXTreated	0.70**	0.53*	0.038	-0.14		0.95***	0.77**	0.88	0.65	
	(0.33)	(0.32)	(0.52)	(0.47)		(0.35)	(0.38)	(0.69)	(0.59)	
MidlineXAdjacent		-0.022		-0.077			1.47***		-0.84**	
		(0.33)		(0.61)			(0.38)		(0.35)	
EndlineXAdjacent		0.39		-0.31			0.30		0.58	
		(0.33)		(0.50)			(0.40)		(0.50)	
Treated					-0.039					0.14
					(0.16)					(0.16)
Constant	-0.53**	-0.27	-0.22	-0.25	0.018	-0.49*	-0.38	0.67	0.57**	-0.065
	(0.21)	(0.25)	(0.58)	(0.55)	(0.073)	(0.29)	(0.29)	(0.40)	(0.29)	(0.072)
Observations	76	159	76	165	247	78	159	76	165	247
Line clusters	38	81	39	88	25	40	81	39	88	25
p: MLxT=MLxA		0.33		0.41			0.42		0.016	
p: ELxT=ELxA		0.64		0.56			0.13		0.85	

Notes: Table 2.D.4 shows the main treatment effects on the HR index and PR index. Control lines are only included if they are located on a different floor as treatment lines. All indices are based on Anderson (2008). Index components are different for Line operators (LO), Line supervisors/line chiefs (LSLC) and managers (MG) (see Section 2.B.3 in the appendix). Regressions include factory fixed effects and the error term is clustered at the line level. p-values are shown at the bottom of the table for testing the equality of coefficients on treatment and adjacent lines at midline and endline (MLxT=MLxA; ELxT=ELxA). Standard errors are shown in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

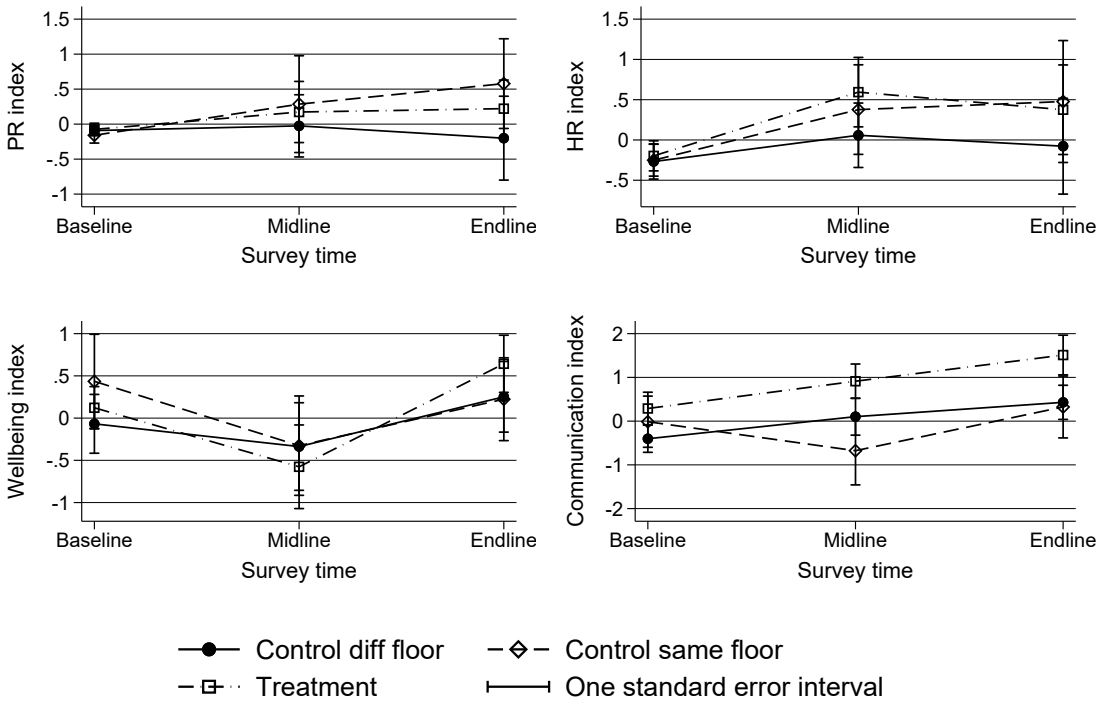


Figure 2.D.17: Dynamics over time for line operators

Notes: Figure 2.D.17 displays standard errors and means of the four main indices (PR, HR, Wellbeing and Communication) for line operators relative to the timing of the training programme by survey group. We differentiate among three survey groups: Operators on treatment lines, operators on control lines on the same floor as treatment lines and operators on control lines on different floors than treatment lines.

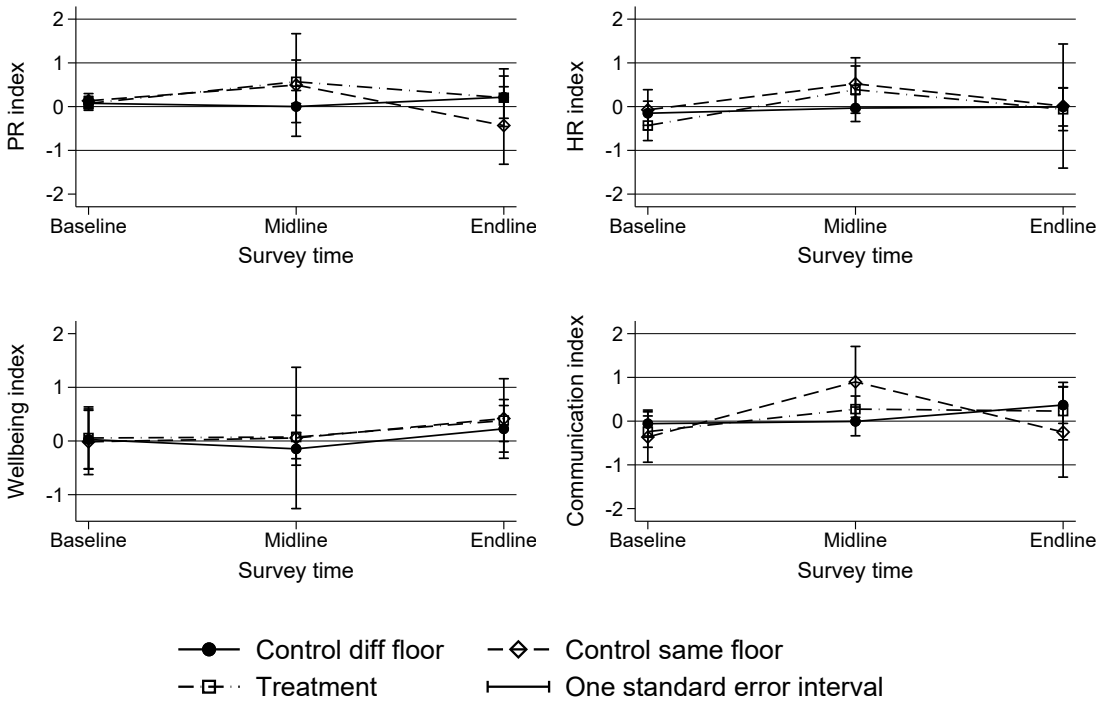


Figure 2.D.18: Dynamics over time for line supervisors/line chiefs

Notes: Figure 2.B.10 displays standard errors and means of the four main indices (PR, HR, Well-being and Communication) for line supervisors/line chiefs relative to the timing of the training programme by survey group. We differentiate among three survey groups: Line supervisors/line chiefs on treatment lines, line supervisors/line chiefs on control lines on the same floor as treatment lines and line supervisors/line chiefs on control lines on different floors than treatment lines.

TABLE 2.D.5. SUB-INDICES FOR LINE OPERATORS: ALL CONTROL LINES

	HR Index						**PR Index**		**Wellb Index**		**Comm Index**	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Induction	Buddy	Leave	WorkRep	Promotion	Bonus	Up-Skill	SkillTest	Feeling	Involve	Practices	Know
Endline	-0.056	0.075	-0.60***	-0.099	0.35	-0.051	0.11	-0.066	0.42*	0.21	-0.026	0.077***
	(0.31)	(0.30)	(0.23)	(0.26)	(0.23)	(0.28)	(0.29)	(0.19)	(0.24)	(0.24)	(0.22)	(0.022)
MidlineXTreated	0.36	1.32***	0.11	-0.10	-0.27	0.0035	-0.035	0.19	-0.20	-0.11	0.86***	0.068***
	(0.31)	(0.30)	(0.24)	(0.28)	(0.30)	(0.26)	(0.25)	(0.24)	(0.26)	(0.28)	(0.30)	(0.025)
EndlineXTreated	-0.020	0.51**	0.54**	0.38	-0.16	-0.042	0.33	-0.034	0.46**	0.098	1.32***	0.033
	(0.35)	(0.24)	(0.22)	(0.24)	(0.27)	(0.24)	(0.32)	(0.23)	(0.22)	(0.23)	(0.28)	(0.032)
MidlineXAdjacent	0.37	0.13	0.23	0.091	0.093	0.22	-0.61**	-0.026	-0.39	0.41	0.86**	0.10***
	(0.38)	(0.29)	(0.23)	(0.26)	(0.30)	(0.24)	(0.25)	(0.25)	(0.24)	(0.28)	(0.33)	(0.024)
EndlineXAdjacent	0.14	-0.081	0.38	0.15	0.16	-0.045	0.15	-0.19	0.25	0.11	0.71**	0.016
	(0.33)	(0.24)	(0.23)	(0.25)	(0.28)	(0.22)	(0.37)	(0.23)	(0.28)	(0.27)	(0.30)	(0.029)
Constant	-0.078	-0.045	0.28	0.0013	0.024	0.070	0.0021	0.010	-0.32*	-0.051	0.063	0.29***
	(0.20)	(0.19)	(0.18)	(0.17)	(0.18)	(0.15)	(0.15)	(0.16)	(0.17)	(0.20)	(0.19)	(0.016)
Observations	111	193	209	209	209	209	209	209	209	209	209	209
Line clusters	72	103	106	106	106	106	106	106	106	106	106	106
p: MLxT=MLxA	0.99	0.00026	0.58	0.51	0.27	0.44	0.047	0.41	0.47	0.052	0.99	0.19
p: ELxT=ELxA	0.68	0.013	0.47	0.32	0.23	0.99	0.59	0.52	0.43	0.96	0.070	0.60

Notes: Table 2.D.5 shows the OLS estimates of treatments effects on each sub-index of the HR, PR, Wellbeing and Communication index for line operators as illustrated in figure 2.1. All indices are constructed using Anderson's (2008) variance-covariance weighted index. A detailed description of each sub-index can be found in figure 2.B.7. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

TABLE 2.D.6. SUB-INDICES FOR LINE SUPERVISORS AND LINE CHIEFS: ALL CONTROL LINES

	HR Index			**PR Index**					**Wellb Index**		**Comm Index**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
	Leave	JobDesc	Bonus	Up-Skill	SkillTest	Quality	Practices	Tools	Feeling	Involve	Practices	Know	Comm IE
Endline	-0.37	0.045	-0.055	-0.95***	-0.10	0.099	0.90***	-0.23	0.66***	0.11	0.54**	-0.34*	-0.044
	(0.22)	(0.26)	(0.27)	(0.21)	(0.17)	(0.21)	(0.26)	(0.31)	(0.25)	(0.27)	(0.24)	(0.18)	(0.23)
MidlineXTreated	-0.074	0.25	-0.070	0.23	0.0086	0.72***	0.26	0.53	-0.27	0.38*	0.29	-0.26	0.12
	(0.23)	(0.24)	(0.23)	(0.33)	(0.20)	(0.25)	(0.27)	(0.52)	(0.36)	(0.23)	(0.23)	(0.19)	(0.27)
EndlineXTreated	0.17	-0.16	-0.011	0.0031	0.40	0.050	-0.30	0.49	-0.27	0.43*	0.16	0.039	0.30
	(0.26)	(0.28)	(0.26)	(0.27)	(0.27)	(0.29)	(0.36)	(0.46)	(0.33)	(0.22)	(0.33)	(0.30)	(0.27)
MidlineXAdjacent	-0.46**	-0.43*	-0.034	-0.51**	-0.20	-0.35	-0.093	-0.40	-0.27	0.084	-0.33*	-0.34	-0.32
	(0.22)	(0.22)	(0.22)	(0.24)	(0.19)	(0.26)	(0.23)	(0.31)	(0.28)	(0.25)	(0.19)	(0.23)	(0.22)
EndlineXAdjacent	-0.092	-0.18	0.12	-0.34**	-0.081	-0.50**	-0.24	-0.073	0.057	-0.058	0.31	0.071	0.13
	(0.25)	(0.23)	(0.21)	(0.15)	(0.24)	(0.25)	(0.24)	(0.33)	(0.29)	(0.28)	(0.30)	(0.18)	(0.27)
Constant	0.24**	0.19	0.031	0.49***	0.14	0.20	-0.55***	0.20	-0.25	-0.053	-0.22*	0.40***	0.069
	(0.12)	(0.13)	(0.14)	(0.16)	(0.12)	(0.15)	(0.17)	(0.19)	(0.18)	(0.19)	(0.12)	(0.14)	(0.13)
Observations	219	219	219	219	219	219	219	219	219	218	219	219	219
Line clusters	116	116	116	116	116	116	116	116	116	116	116	116	116
p: MLxT=MLxA	0.15	0.010	0.88	0.029	0.30	0.00026	0.19	0.078	0.98	0.11	0.0059	0.71	0.12
p: ELxT=ELxA	0.24	0.92	0.58	0.14	0.051	0.026	0.85	0.19	0.30	0.045	0.62	0.91	0.49

Notes: Table 2.D.6 shows the OLS estimates of treatments effects on each sub-index of the HR, PR, Wellbeing and Communication index for line supervisors / line chiefs as illustrated in figure 2.1. All indices are constructed using Anderson's (2008) variance-covariance weighted index. A detailed description of each sub-index can be found in figure 2.B.8. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

TABLE 2.D.7. SUB-INDICES FOR MANAGERS: ALL CONTROL LINES

	HR Index					**PR Index**				**Wellb Index**		**Comm Index**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Buddy	Leave	Promotion	Exit	Bonus	Up-Skill	Test	Quality	Practices	Feeling	Feeling LO	Practices
Treated	0.64***	0.075	-0.16	-0.023	-0.13	-0.013	0.20	-0.68***	-0.20	-0.072	0.0074	0.14
	(0.11)	(0.13)	(0.12)	(0.13)	(0.10)	(0.096)	(0.14)	(0.11)	(0.16)	(0.16)	(0.15)	(0.16)
Constant	-0.29***	-0.035	0.073	0.010	0.060	0.0060	-0.093	0.31***	0.091	0.033	-0.0034	-0.065
	(0.051)	(0.060)	(0.056)	(0.059)	(0.048)	(0.044)	(0.066)	(0.052)	(0.075)	(0.074)	(0.068)	(0.072)
Observations	247	247	247	247	246	247	247	247	247	247	247	247
Line clusters	25	25	25	25	25	25	25	25	25	25	25	25

Notes: Table 2.D.7 shows the OLS estimates of treatments effects on each sub-index of the HR, PR, Wellbeing and Communication index for managers as illustrated in figure 2.1. All indices are constructed using Anderson's (2008) variance-covariance weighted index. A detailed description of each sub-index can be found in figure 2.B.9. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

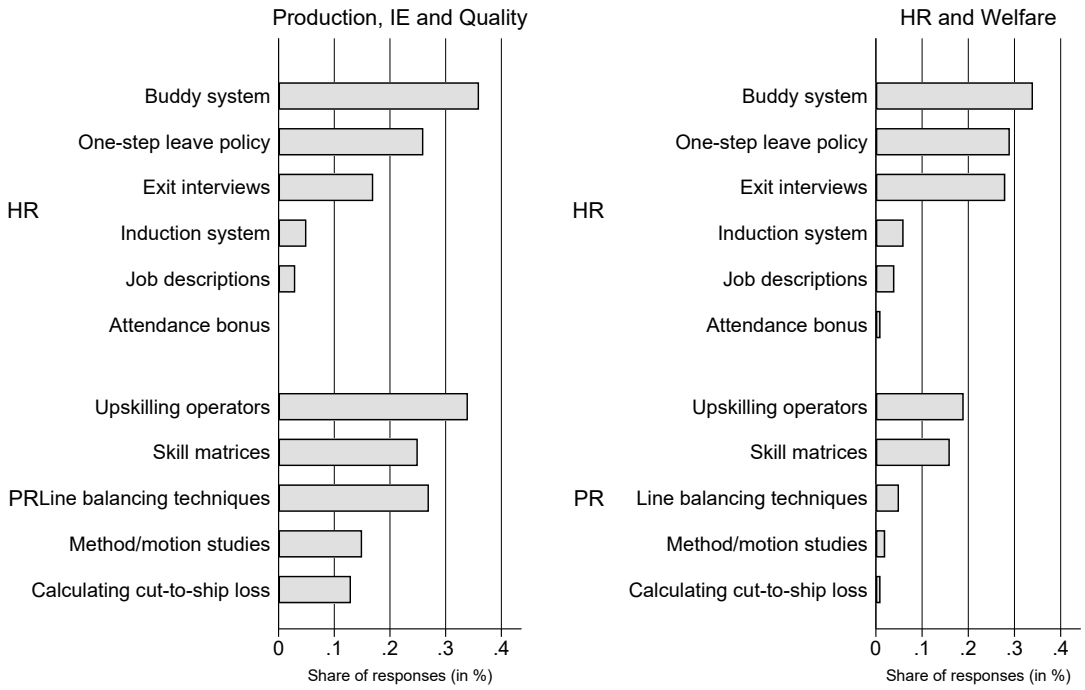


Figure 2.D.19: Managers' most important perceived changes in policies

Notes: Figure 2.D.19 displays the percentage share of responses of managers to the question "What are the most important changes in policies or processes that have resulted from BBW?". Managers could give several answer choices. The left graph displays the share of responses of managers in the production, IE or quality department per answer choice. The right graph displays the share of responses of managers in the HR or welfare department. Responses are clustered into HR- and PR-related topics.

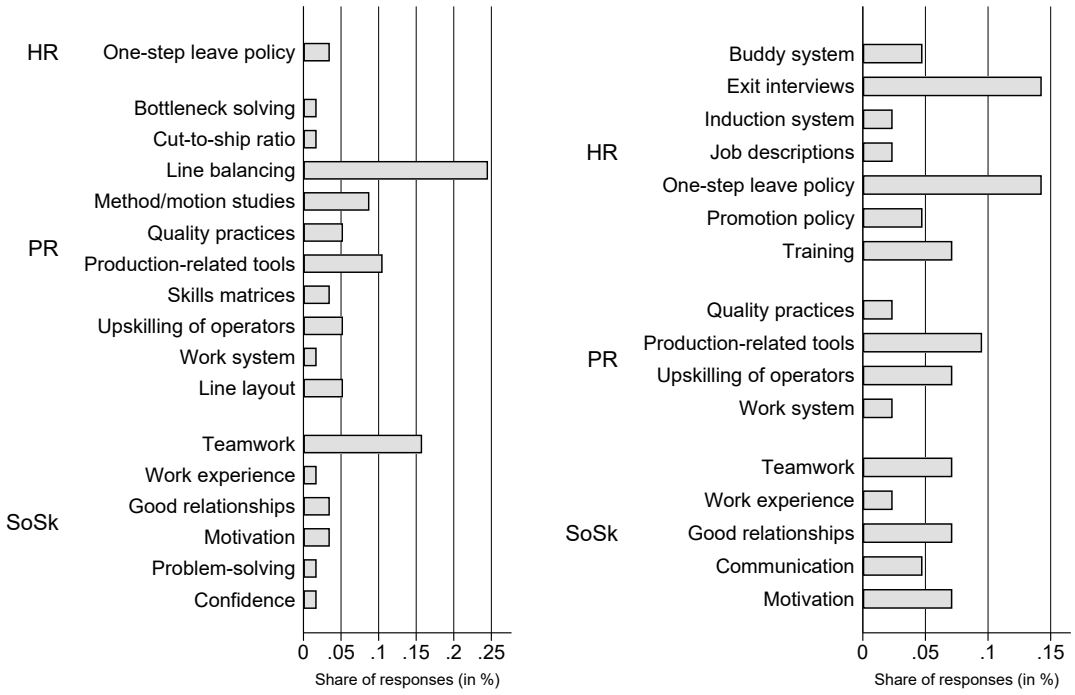


Figure 2.D.20: Managers' most important perceived changes in own job

Notes: Figure 2.D.20 displays the percentage share of responses of managers to the question "In your own job, what has changed the most as a result of BBW?". Managers could only give one answer choice. The left graph displays answer choices of managers in the production, IE or quality department. The right graph displays answer choices of managers in the HR or welfare department. Responses are clustered into HR-, PR- and communication/soft skills-related topics.

TABLE 2.D.8. SURVEY RESULTS OF THE BUDDY SYSTEM
(RESPONDENTS: LINE OPERATORS)

	All		Only new LO (tenure < 60 days)			Only old LO	Only mentors	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Index	Paired	1st talk to buddy	Mentor useful	Have mentor	Be mentor	Questions	Teach. ability
Endline	0.075	-0.055	0.14	-0.20	0.26	-0.0081	-0.092	1.22**
	(0.30)	(0.20)	(0.26)	(0.22)	(0.33)	(0.28)	(1.12)	(0.54)
MidlineXTreated	1.32***	-0.40	0.99**	0.53	0.36	0.96***	0.64	0.75
	(0.30)	(0.28)	(0.40)	(0.35)	(0.32)	(0.25)	(0.64)	(0.45)
EndlineXTreated	0.51**	0.014	0.28	0.67*	-0.088	0.48**	0.36	-0.73*
	(0.24)	(0.27)	(0.34)	(0.36)	(0.36)	(0.22)	(0.64)	(0.42)
MidlineXAdjacent	0.13	0.094	-0.12	0.16	0.090	0.15	0.35	-0.15
	(0.29)	(0.24)	(0.27)	(0.45)	(0.40)	(0.27)	(0.57)	(0.61)
EndlineXAdjacent	-0.081	0.091	0.090	0.11	0.032	0.039	0.48	-0.73
	(0.24)	(0.45)	(0.27)	(0.31)	(0.46)	(0.25)	(0.73)	(0.70)
Observations	193	111	111	111	97	177	72	76
Line clusters	103	72	72	72	67	100	53	53

Notes: Table 2.D.8 presents the treatment effects after the midline survey (MLxT) and the endline survey (ELxT) on the buddy system. Column 1 shows the overall Anderson index, columns 2 to 8 the individual variables included in the index. Survey questions were asked to line operators only, where columns 2 to 5 only include new operators (with tenure <60 days), column 6 only old line operators (tenure >60 days) and columns 7 and 8 only mentors. The variables are defined as follows. "Paired": Operator was sitting next to experienced worker during first months at factory. "1st talk to buddy": Buddy is first person to talk to if worker has work-related grievance. "Mentor useful": Buddy is most helpful during first weeks at the factory. "Have mentor": Buddy exists on the line. "Be mentor": Worker has ever been assigned as buddy. "Questions": Frequency of being asked questions by new operators. "Teaching ability": Assessment of teaching and mentoring ability. Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01

TABLE 2.D.9. SURVEY RESULTS OF THE ONE-STEP LEAVE POLICY
(RESPONDENTS: LINE OPERATORS)

	(1)	(2)	(3)
	Index	One-step	Time away
Endline	-0.60***	0.037	-0.86***
	(0.23)	(0.19)	(0.23)
MidlineXTreated	0.11	0.36	-0.14
	(0.24)	(0.25)	(0.22)
EndlineXTreated	0.54**	0.26	0.53**
	(0.22)	(0.22)	(0.21)
MidlineXAdjacent	0.23	0.28	0.094
	(0.23)	(0.26)	(0.20)
EndlineXAdjacent	0.38	0.19	0.36
	(0.23)	(0.22)	(0.25)
Observations	209	209	209
Line clusters	106	106	106

Notes: Table 2.D.9 presents the treatment effects after the midline survey (MLxT) and the endline survey (ELxT) on the one-step leave policy. Column 1 shows the overall Anderson index, columns 2 and 3 the individual variables included in the index. The variables are defined as follows. "One-step": Only one person approached to request leave. "Time away": Time spend away to request leave. Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01

TABLE 2.D.10. SURVEY RESULTS OF OPERATORS' FEELINGS
(RESPONDENTS: LINE OPERATORS)

	(1)	(2)	(3)	(4)
	Index	Conflict with MG	Self assessment	Feelings of others
Endline	0.42*	-0.30	0.67***	0.73***
	(0.24)	(0.22)	(0.21)	(0.25)
MidlineXTreated	-0.20	0.13	-0.31*	-0.32
	(0.26)	(0.24)	(0.18)	(0.20)
EndlineXTreated	0.46**	0.29	0.39	0.056
	(0.22)	(0.18)	(0.25)	(0.24)
MidlineXAdjacent	-0.39	-0.19	-0.11	-0.35*
	(0.24)	(0.22)	(0.19)	(0.20)
EndlineXAdjacent	0.25	0.22	0.23	-0.080
	(0.28)	(0.23)	(0.25)	(0.28)
Observations	209	209	209	209
Line clusters	106	106	106	106

Notes: Table 2.D.10 presents the treatment effects after the midline survey (MLxT) and the endline survey (ELxT) on operators' feelings. Column 1 shows the overall Anderson index, columns 2 to 4 the individual variables included in the index. Survey questions were asked to line operators only. The variables are defined as follows. "Conflict with MG": No perceived conflict between workers and managers. "Self assessment": Assessment of own feelings towards job. "Feelings of others": Assessment of feelings of others towards job. Standard errors in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

TABLE 2.D.11. SURVEY RESULTS OF COMMUNICATION PRACTICES (RESPONDENTS: LINE OPERATORS)

	(1) Index	(2) Initiate meetings	(3) Approach supervisors	(4) Discuss suggestions
Endline	-0.026	-0.15	0.0073	0.099
	(0.22)	(0.19)	(0.18)	(0.17)
MidlineXTreated	0.86***	0.49**	0.47***	0.37
	(0.30)	(0.21)	(0.17)	(0.24)
EndlineXTreated	1.32***	0.80***	0.91***	0.44**
	(0.28)	(0.20)	(0.19)	(0.21)
MidlineXAdjacent	0.86**	0.56**	0.33	0.37
	(0.33)	(0.22)	(0.24)	(0.24)
EndlineXAdjacent	0.71**	0.47**	0.38*	0.27
	(0.30)	(0.23)	(0.22)	(0.20)
Observations	209	209	209	209
Line clusters	106	106	106	106

Notes: Table 2.D.11 presents the treatment effects after the midline survey (MLxT) and the endline survey (ELxT) on communication practices. Column 1 shows the overall Anderson index, columns 2 to 4 the individual variables included in the index. Survey questions were asked to line operators only. The variables are defined as follows. "Initiate meetings": Variety of topics for which worker initiate meetings with immediate superior (production capabilities, skills and/or work grievances). "Approach supervisor": Frequency of approaching supervisors to tell them about issues that affect workers. "Discuss suggestions": Frequency of talking to other operators about how to make things better at the factory. Standard errors in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

TABLE 2.D.12. SURVEY RESULTS OF SUPERVISORS/LINE CHIEF KNOWLEDGE (RESPONDENTS: LINE OPERATORS)

	(1) Index	(2) 1st position	(3) Married	(4) Length	(5) # Child.	(6) # Boys	(7) # Girls	(8) District	(9) Division	(10) Years	(11) Months
Endline	0.077***	-0.14	0.34	0.16	-0.065	-0.24	0.33	1.18***	-0.17	-0.064	0.34
	(0.022)	(0.16)	(0.23)	(0.34)	(0.27)	(0.24)	(0.31)	(0.21)	(0.20)	(0.19)	(0.24)
MidlineXTreated	0.068***	-0.038	0.43*	0.028	0.36	0.46*	0.52	0.029	-0.14	0.17	0.30
	(0.025)	(0.23)	(0.25)	(0.38)	(0.27)	(0.24)	(0.31)	(0.10)	(0.21)	(0.24)	(0.22)
EndlineXTreated	0.033	0.32	0.33	0.079	0.012	-0.022	-0.45	0.32	0.50**	0.16	-0.20
	(0.032)	(0.22)	(0.26)	(0.37)	(0.29)	(0.37)	(0.32)	(0.31)	(0.22)	(0.23)	(0.27)
MidlineXAdjacent	0.10***	0.33	0.88***	0.58*	0.16	0.097	0.32	-0.050	0.41*	-0.13	0.014
	(0.024)	(0.23)	(0.23)	(0.34)	(0.30)	(0.30)	(0.32)	(0.097)	(0.23)	(0.23)	(0.19)
EndlineXAdjacent	0.016	0.18	0.072	-0.57*	0.17	0.16	-0.23	0.062	0.20	0.37	-0.052
	(0.029)	(0.23)	(0.29)	(0.34)	(0.25)	(0.37)	(0.32)	(0.29)	(0.23)	(0.25)	(0.29)
Observations	209	209	209	113	158	139	139	209	209	209	209
Line clusters	106	106	106	80	100	87	87	106	106	106	106

Notes: Table 2.D.12 presents the treatment effects after the midline survey (MLxT) and the endline survey (ELxT) on the knowledge operators have about their line supervisors and line chiefs. Column 1 shows the overall index based on Katz et al. (2007), columns 2 to 11 the individual standardized variables which are aggregated into the index. Survey questions were asked to line operators only. The aim was to figure out how much private information operators know about the line supervisor/line chief (LS/LC) they have worked most with on the current line, in particular about their first position in the garment industry(column 1), marital status (column 2), length of marriage (column 3), number of children and if >0: number of boys and girls (column 5 to 7), home district and division (column 8 and 9) and tenure in the factory in years and months (column 10 and 11). Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01

TABLE 2.D.13. SURVEY RESULTS OF QUALITY PRACTICES (RESPONDENTS: LINE SUPERVISORS/ LINE CHIEFS)

	(1) Index	(2) Quality checks	(3) What changes	(4) Inspection	(5) File checks	(6) Flags
Endline	0.099	-0.22	-0.33*	0.21	0.19	0.064
	(0.21)	(0.21)	(0.19)	(0.17)	(0.22)	(0.20)
MidlineXTreated	0.72***	0.57**	0.78**	0.38	0.42*	0.090
	(0.25)	(0.25)	(0.30)	(0.23)	(0.23)	(0.26)
EndlineXTreated	0.050	0.049	0.30	-0.11	0.091	0.11
	(0.29)	(0.26)	(0.27)	(0.23)	(0.25)	(0.25)
MidlineXAdjacent	-0.35	-0.18	-0.30	-0.10	-0.20	-0.26
	(0.26)	(0.22)	(0.20)	(0.23)	(0.21)	(0.22)
EndlineXAdjacent	-0.50**	-0.59***	-0.43***	-0.30	0.10	-0.16
	(0.25)	(0.21)	(0.16)	(0.21)	(0.23)	(0.21)
Observations	219	219	219	219	219	219
Line clusters	116	116	116	116	116	116

Notes: Table 2.D.13 presents the treatment effects after the midline survey (MLxT) and the endline survey (ELxT) on quality practices. Column 1 shows the overall Anderson index, and columns 2 to 6 the individual standardized variables included in the index. "Quality checks": Change to quality checks in last six months. "What changes": Number of changes. "Inspection": Lines uses endline inspection format. "File checks": Line maintains a file for tracking quantity of garments checked. "Flags": Flags used as a method/aid in quality inspections. Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01

TABLE 2.D.14. SURVEY RESULTS OF BUDDY SYSTEM
(RESPONDENTS: MANAGERS)

	(1) Index	(2) Buddy in line	(3) Number of buddy lines	(4) Duration buddy system	(5) Buddy issues	(6) Buddy issues frequency
Treated	0.64***	-0.16	0.014	1.71***	0.092	-0.097
	(0.11)	(0.12)	(0.072)	(0.11)	(0.12)	(0.18)
Observations	247	247	226	223	247	133
Line clusters	25	25	25	25	25	25

Notes: Table 2.D.14 presents the treatment effects on the buddy system. Column 1 shows the overall Anderson index, and columns 2 to 6 the individual standardized variables included in the index. "Buddy in line": Indicator for buddy on pilot lines. "Number of buddy lines": Number of lines with buddy. "Duration buddy system": Amount of time buddy system has been active. "Buddy issues": Indicator for buddies approaching managers about worker issues. "Buddy issues frequency": Frequency of approaches by buddies to managers about worker issues. Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01

TABLE 2.D.15. SURVEY RESULTS OF QUALITY PRACTICES
(RESPONDENTS: MANAGERS)

	(1) Index	(2) New system	(3) Additions	(4) Zero defect operators	(5) Zero defect lines
Treated	-0.68***	-0.55***	-0.79***	-0.20	0.42
	(0.11)	(0.13)	(0.12)	(0.17)	(0.38)
Observations	247	247	247	108	99
Line clusters	25	25	25	22	21

Notes: Table 2.D.15 presents the treatment effects on quality practices. Column 1 shows the overall Anderson index, and columns 2 to 5 the individual standardized variables included in the index. "New system": Indicator for implementation of new quality system in past 6 months. "Additions": Number of new practices/systems added. "Zero defect operators": Number of zero defect operators on pilot lines. "Zero defect lines": Number of zero defect lines in factory. Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01

2.E Supplementary figures and tables: Firm outcomes

All regressions in this Section include time fixed effects to control for seasonal variations. We use weekly time dummies for the PR index and corresponding index components, and monthly time dummies for the HR index and corresponding index components. Factory fixed effects are also included to control for systematic differences across factories. The error term is clustered at the line level. At the bottom of each table, we report false discovery rate-adjusted q-values for HR-related and PR-related outcome families. We use the 'sharpened q value' approach based on the Benjamini-Hochberg procedure Benjamini et al. (2006) to adjust for multiple outcome variables. For simplicity, tables do not report coefficients on the $PostX$ variables for $X = (1, 4)$. The baseline value of the dependent variable is also not shown.

TABLE 2.E.16. OVERALL EFFECTS ON HR AND PR VARIABLES: ALL CONTROL LINES (COLUMNS 3 AND 4 IN SUMMARY TABLE 2.9)

	Index		**HR Variables**				**PR Variables**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	HR	**PR**	Exit	Turnover	Absenteeism	Promotion	Efficiency	Alterations	Rejects
TreatedXPost1	0.207	0.002	-0.030	-0.011	-0.001	0.004	0.045*	0.004	0.000
	(0.209)	(0.230)	(0.078)	(0.011)	(0.010)	(0.005)	(0.023)	(0.013)	(0.001)
			{1.000}	{1.000}	{1.000}	{1.000}	{0.195}	{1.000}	{1.000}
TreatedXPost2	0.447	-0.004	-0.133	-0.036	0.007	0.005	0.050	0.008	0.000
	(0.439)	(0.148)	(0.082)	(0.026)	(0.008)	(0.025)	(0.039)	(0.011)	(0.001)
			{0.548}	{0.548}	{0.548}	{0.726}	{1.000}	{1.000}	{1.000}
TreatedXPost3	0.396	0.005	0.105	0.003	0.000	0.034	0.006	0.002	-0.000
	(0.493)	(0.164)	(0.159)	(0.016)	(0.008)	(0.027)	(0.049)	(0.011)	(0.000)
			{1.000}	{1.000}	{1.000}	{1.000}	{1.000}	{1.000}	{1.000}
TreatedXPost4	-0.202	0.088	0.012	0.009	0.007	-0.006	0.016	0.002	-0.001**
	(0.229)	(0.121)	(0.056)	(0.012)	(0.007)	(0.005)	(0.028)	(0.010)	(0.000)
			{1.000}	{1.000}	{1.000}	{1.000}	{1.000}	{1.000}	{0.057}
Observations	200	1631	200	200	185	200	751	1630	1630
Clusters	35	52	35	35	32	35	30	52	52
R^2	0.34	0.33	0.28	0.33	0.52	0.21	0.51	0.30	0.32
Control mean	0.05	0.02	0.24	0.06	0.04	0.01	0.37	0.06	0.00
Factory FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes: Table 2.E.16 shows treatment effects on HR and PR indices in column one and two, and individual index components (exit, turnover, absenteeism and promotion rates for HR index, and efficiency, alteration rates and reject rates for PR index). Regression results are shown for factories of type "AB", which have control lines on different floors. Columns one and two refer to columns 3 and 4 in the summary Table 2.9. Regressions include time fixed effects to control for seasonal variations. We use weekly time dummies for the PR index and corresponding index components, and monthly time dummies for the HR index and corresponding index components. Factory fixed effects are also included to control for systematic differences across factories. The error term is clustered at the line level. False discovery rate-adjusted p-values, also known as q-values, for HR-related and PR-related outcome families were used to adjust for multiple outcome variables. They are shown in curly brackets. We use the 'sharpened q value' approach based on the Benjamini-Hochberg procedure (Benjamini et al., 2006). For simplicity, the regression table does not report coefficients on the $PostX$ variables for $X = (1, 4)$, and the baseline value of the dependent variables. Standard errors are shown in round parentheses. * p<0.1, ** p<0.05, *** p<0.01

TABLE 2.E.17. OVERALL EFFECTS ON HR AND PR VARIABLES: DIFF. CONTROL LINES (COLUMNS 5 AND 6 IN SUMMARY TABLE 2.9)

	Index		**HR Variables**				**PR Variables**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	HR	**PR**	Exit	Turnover	Absenteeism	Promotion	Efficiency	Alterations	Rejects
TreatedXPost1	0.282	0.016	-0.052	-0.012	-0.001	0.005	0.045*	0.003	0.000
	(0.219)	(0.237)	(0.082)	(0.012)	(0.011)	(0.005)	(0.025)	(0.014)	(0.001)
			{1.000}	{1.000}	{1.000}	{1.000}	{0.322}	{1.000}	{1.000}
TreatedXPost2	0.493	-0.057	-0.136	-0.045	0.003	0.003	0.055	0.010	0.000
	(0.450)	(0.154)	(0.087)	(0.029)	(0.008)	(0.026)	(0.040)	(0.012)	(0.001)
			{0.392}	{0.392}	{0.838}	{0.838}	{0.965}	{0.965}	{0.965}
TreatedXPost3	0.445	-0.009	0.123	0.002	-0.002	0.036	0.006	0.001	-0.000
	(0.513)	(0.168)	(0.160)	(0.017)	(0.009)	(0.027)	(0.049)	(0.012)	(0.000)
			{1.000}	{1.000}	{1.000}	{1.000}	{1.000}	{1.000}	{1.000}
TreatedXPost4	-0.166	0.070	0.002	0.005	0.008	-0.006	0.020	0.003	-0.001**
	(0.234)	(0.123)	(0.058)	(0.013)	(0.008)	(0.006)	(0.027)	(0.010)	(0.000)
			{1.000}	{1.000}	{1.000}	{1.000}	{0.843}	{0.975}	{0.132}
Observations	188	1511	188	188	173	188	710	1510	1510
Clusters	33	48	33	33	30	33	27	48	48
R^2	0.32	0.32	0.29	0.34	0.49	0.21	0.51	0.30	0.29
Control mean	0.07	0.02	0.25	0.07	0.04	0.01	0.37	0.06	0.00
Factory FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes: Table 2.E.16 shows treatment effects on HR and PR indices in column one and two, and individual index components (exit, turnover, absenteeism and promotion rates for HR index, and efficiency, alteration rates and reject rates for PR index). Regression results are shown for factories of type "AB", which have control lines on different floors, but excludes control floors on the same floor as of treatment lines. Columns one and two refer to columns 5 and 6 in the summary Table 2.9. Regressions include time fixed effects to control for seasonal variations. We use weekly time dummies for the PR index and corresponding index components, and monthly time dummies for the HR index and corresponding index components. Factory fixed effects are also included to control for systematic differences across factories. The error term is clustered at the line level. False discovery rate-adjusted p-values, also known as q-values, for HR-related and PR-related outcome families were used to adjust for multiple outcome variables. They are shown in curly brackets. We use the 'sharpened q value' approach based on the Benjamini-Hochberg procedure (Benjamini et al., 2006). For simplicity, the regression table does not report coefficients on the $PostX$ variables for $X = (1, 4)$, and the baseline value of the dependent variables. Standard errors are shown in round parentheses. * p<0.1, ** p<0.05, *** p<0.01

TABLE 2.E.18. SPILLOVER EFFECTS: ADJACENT LINES (ABC) (COLUMNS 7 AND 8 IN SUMMARY TABLE 2.9)

	Index		**HR Variables**				**PR Variables**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	HR	**PR**	Exit	Turnover	Absenteeism	Promotion	Efficiency	Alterations	Rejects
TreatedXPost1	0.450*	0.037	-0.076	-0.031*	-0.003	0.011	0.041**	-0.001	0.000
	(0.234)	(0.185)	(0.060)	(0.016)	(0.007)	(0.014)	(0.020)	(0.011)	(0.001)
			{0.449}	{0.303}	{0.704}	{0.704}	{0.156}	{1.000}	{1.000}
TreatedXPost2	0.464**	-0.000	-0.163**	-0.037*	0.000	0.002	0.059*	0.005	0.000
	(0.206)	(0.133)	(0.069)	(0.021)	(0.005)	(0.016)	(0.035)	(0.009)	(0.001)
			{0.089}	{0.148}	{0.947}	{0.947}	{0.388}	{1.000}	{1.000}
TreatedXPost3	0.018	0.010	0.173	-0.005	0.000	0.029	0.024	0.000	-0.000
	(0.400)	(0.139)	(0.165)	(0.016)	(0.007)	(0.022)	(0.041)	(0.009)	(0.000)
			{1.000}	{1.000}	{1.000}	{1.000}	{1.000}	{1.000}	{1.000}
TreatedXPost4	-0.116	0.093	-0.002	0.011	0.001	-0.006	0.025	-0.001	-0.001**
	(0.150)	(0.105)	(0.040)	(0.013)	(0.006)	(0.005)	(0.030)	(0.008)	(0.000)
			{1.000}	{1.000}	{1.000}	{1.000}	{0.718}	{1.000}	{0.105}
AdjacentXPost1	0.159	0.062	0.009	-0.017	-0.001	0.005	0.029	-0.002	0.000
	(0.191)	(0.150)	(0.068)	(0.015)	(0.006)	(0.008)	(0.020)	(0.009)	(0.000)
			{1.000}	{1.000}	{1.000}	{1.000}	{0.811}	{1.000}	{1.000}
AdjacentXPost2	-0.014	0.096	0.108	-0.045**	0.004	0.002	0.070**	-0.003	-0.000
	(0.242)	(0.100)	(0.087)	(0.020)	(0.006)	(0.016)	(0.030)	(0.006)	(0.000)
			{0.495}	{0.120}	{0.791}	{0.823}	{0.079}	{1.000}	{1.000}
AdjacentXPost3	-0.348	-0.070	0.033	0.081	0.004	0.020	0.009	0.005	-0.000
	(0.564)	(0.132)	(0.132)	(0.051)	(0.009)	(0.016)	(0.033)	(0.009)	(0.000)
			{0.890}	{0.848}	{0.890}	{0.848}	{1.000}	{1.000}	{1.000}
AdjacentXPost4	0.006	0.059	-0.034	0.008	-0.001	-0.003	-0.010	-0.004	-0.000
	(0.114)	(0.094)	(0.041)	(0.011)	(0.004)	(0.005)	(0.024)	(0.006)	(0.000)
			{1.000}	{1.000}	{1.000}	{1.000}	{1.000}	{1.000}	{1.000}
Observations	469	3397	469	469	383	469	1406	3396	3396
Clusters	83	112	83	83	67	83	56	112	112
R^2	0.29	0.31	0.23	0.25	0.51	0.23	0.50	0.27	0.34
Control mean	0.04	-0.05	0.22	0.06	0.03	0.01	0.41	0.06	0.00
Factory FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes: Table 2.E.18 shows treatment effects on HR and PR indices in column one and two, which refer to columns 7 and 8 in the summary Table 2.9, and individual index components (exit, turnover, absenteeism and promotion rates for HR index, and efficiency, alteration rates and reject rates for PR index). Regressions include time fixed effects to control for seasonal variations. We use weekly time dummies for the PR index and corresponding index components, and monthly time dummies for the HR index and corresponding index components. Factory fixed effects are also included to control for systematic differences across factories. The error term is clustered at the line level. False discovery rate-adjusted p-values, also known as q-values, for HR-related and PR-related outcome families were used to adjust for multiple outcome variables. They are shown in curly brackets. We use the 'sharpened q value' approach based on the Benjamini-Hochberg procedure (Benjamini et al., 2006). For simplicity, the regression table does not report coefficients on the $PostX$ variables for $X = (1, 4)$, and the baseline value of the dependent variables. Standard errors are shown in round parentheses. * p<0.1, ** p<0.05, *** p<0.01

TABLE 2.E.19. SPILLOVER EFFECTS: ADJACENT LINES (COLUMNS 9 AND 10 IN SUMMARY TABLE 2.9)

	Index		**HR Variables**				**PR Variables**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	HR	**PR**	Exit	Turnover	Absenteeism	Promotion	Efficiency	Alterations	Rejects
TreatedXPost1	0.287	0.016	-0.054	-0.012	-0.002	0.005	0.047**	0.002	0.000
	(0.207)	(0.232)	(0.077)	(0.011)	(0.011)	(0.005)	(0.021)	(0.014)	(0.001)
			{1.000}	{1.000}	{1.000}	{1.000}	{0.093}	{1.000}	{1.000}
TreatedXPost2	0.509	-0.050	-0.143	-0.045	0.003	0.003	0.058	0.009	0.000
	(0.423)	(0.150)	(0.087)	(0.028)	(0.007)	(0.025)	(0.037)	(0.011)	(0.001)
			{0.315}	{0.315}	{0.818}	{0.829}	{0.592}	{0.592}	{0.592}
TreatedXPost3	0.459	-0.004	0.111	0.002	-0.001	0.036	0.008	0.000	-0.000
	(0.525)	(0.166)	(0.155)	(0.016)	(0.008)	(0.027)	(0.046)	(0.012)	(0.000)
			{1.000}	{1.000}	{1.000}	{1.000}	{1.000}	{1.000}	{1.000}
TreatedXPost4	-0.153	0.084	-0.003	0.005	0.007	-0.006	0.029	0.002	-0.001*
	(0.225)	(0.122)	(0.057)	(0.013)	(0.007)	(0.006)	(0.032)	(0.010)	(0.000)
			{1.000}	{1.000}	{1.000}	{1.000}	{0.586}	{1.000}	{0.184}
AdjacentXPost1	-0.022	0.011	0.074	0.004	-0.002	0.004	0.034	0.002	0.000
	(0.219)	(0.189)	(0.091)	(0.012)	(0.009)	(0.004)	(0.021)	(0.012)	(0.001)
			{1.000}	{1.000}	{1.000}	{1.000}	{0.530}	{1.000}	{1.000}
AdjacentXPost2	0.386	0.046	-0.023	-0.048*	0.003	0.008	0.075**	0.000	0.000
	(0.440)	(0.112)	(0.103)	(0.027)	(0.008)	(0.024)	(0.032)	(0.008)	(0.000)
			{1.000}	{0.486}	{1.000}	{1.000}	{0.071}	{1.000}	{1.000}
AdjacentXPost3	0.331	-0.001	-0.021	0.028*	0.003	0.023	0.018	-0.002	-0.000
	(0.531)	(0.146)	(0.130)	(0.015)	(0.009)	(0.019)	(0.040)	(0.010)	(0.000)
			{0.770}	{0.381}	{0.770}	{0.477}	{1.000}	{1.000}	{1.000}
AdjacentXPost4	0.120	0.036	-0.081	-0.006	0.003	-0.003	-0.001	-0.001	-0.000
	(0.144)	(0.107)	(0.048)	(0.010)	(0.005)	(0.006)	(0.026)	(0.007)	(0.000)
			{0.689}	{0.873}	{0.873}	{0.873}	{1.000}	{1.000}	{1.000}
Observations	297	2486	297	297	272	297	1184	2485	2485
Clusters	52	79	52	52	47	52	44	79	79
R^2	0.31	0.32	0.24	0.32	0.48	0.21	0.48	0.31	0.30
Control mean	0.17	-0.01	0.22	0.06	0.03	0.01	0.40	0.06	0.00
Factory FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes: Table 2.E.19 shows treatment effects on HR and PR indices in column one and two, which refer to columns 9 and 10 in the summary Table 2.9, and individual index components (exit, turnover, absenteeism and promotion rates for HR index, and efficiency, alteration rates and reject rates for PR index). Regressions include time fixed effects to control for seasonal variations. We use weekly time dummies for the PR index and corresponding index components, and monthly time dummies for the HR index and corresponding index components. Factory fixed effects are also included to control for systematic differences across factories. The error term is clustered at the line level. False discovery rate-adjusted p-values, also known as q-values, for HR-related and PR-related outcome families were used to adjust for multiple outcome variables. They are shown in curly brackets. We use the 'sharpened q value' approach based on the Benjamini-Hochberg procedure (Benjamini et al., 2006). For simplicity, the regression table does not report coefficients on the $PostX$ variables for $X = (1, 4)$, and the baseline value of the dependent variables. Standard errors are shown in round parentheses. * p<0.1, ** p<0.05, *** p<0.01

TABLE 2.E.20. SPILLOVER EFFECTS: ADJACENT AND OTHER LINES (COLUMNS 11 AND 12 IN SUMMARY TABLE 2.9)

	Index		**HR Variables**				**PR Variables**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	HR	**PR**	Exit	Turnover	Absenteeism	Promotion	Efficiency	Alterations	Rejects
TreatedXPost1	0.204	0.045	-0.086*	-0.010	0.000	0.001	0.035**	0.001	-0.000
	(0.129)	(0.220)	(0.049)	(0.007)	(0.008)	(0.003)	(0.017)	(0.012)	(0.001)
			{0.484}	{0.484}	{0.980}	{0.980}	{0.126}	{1.000}	{1.000}
TreatedXPost2	0.470	-0.120	-0.162***	-0.032*	0.001	0.003	0.018	0.013	0.000
	(0.354)	(0.133)	(0.051)	(0.017)	(0.005)	(0.020)	(0.032)	(0.011)	(0.000)
			{0.007}	{0.089}	{0.811}	{0.811}	{1.000}	{1.000}	{1.000}
TreatedXPost3	0.602	-0.053	0.019	-0.037	0.000	0.027	-0.001	0.003	-0.000
	(0.496)	(0.154)	(0.125)	(0.024)	(0.006)	(0.026)	(0.046)	(0.009)	(0.000)
			{1.000}	{0.971}	{1.000}	{0.971}	{1.000}	{1.000}	{1.000}
TreatedXPost4	-0.054	0.076	0.007	-0.008	0.003	-0.003	0.001	-0.001	-0.001*
	(0.191)	(0.104)	(0.047)	(0.010)	(0.007)	(0.004)	(0.025)	(0.010)	(0.000)
			{1.000}	{1.000}	{1.000}	{1.000}	{1.000}	{1.000}	{0.275}
AdjacentXPost1	-0.128	0.055	0.054	0.006	0.001	0.000	0.015	-0.001	-0.000
	(0.155)	(0.162)	(0.072)	(0.008)	(0.006)	(0.003)	(0.018)	(0.010)	(0.001)
			{1.000}	{1.000}	{1.000}	{1.000}	{1.000}	{1.000}	{1.000}
AdjacentXPost2	0.337	-0.054	-0.041	-0.037**	0.001	0.008	0.031	0.003	0.001*
	(0.382)	(0.084)	(0.075)	(0.014)	(0.005)	(0.020)	(0.024)	(0.006)	(0.000)
			{1.000}	{0.051}	{1.000}	{1.000}	{0.237}	{0.402}	{0.235}
AdjacentXPost3	0.410	-0.049	-0.094	-0.012	0.004	0.015	0.013	0.002	-0.000
	(0.479)	(0.128)	(0.091)	(0.024)	(0.008)	(0.018)	(0.033)	(0.008)	(0.000)
			{1.000}	{1.000}	{1.000}	{1.000}	{1.000}	{1.000}	{1.000}
AdjacentXPost4	0.195*	0.042	-0.062*	-0.020***	-0.001	-0.000	-0.008	-0.004	-0.000
	(0.104)	(0.084)	(0.036)	(0.007)	(0.003)	(0.004)	(0.028)	(0.006)	(0.000)
			{0.147}	{0.009}	{0.898}	{0.898}	{1.000}	{1.000}	{1.000}
OtherXPost1	0.144	0.176	-0.067	-0.012*	-0.001	-0.001	0.004	-0.011*	-0.000
	(0.109)	(0.122)	(0.045)	(0.007)	(0.004)	(0.003)	(0.013)	(0.006)	(0.001)
			{0.396}	{0.396}	{0.750}	{0.657}	{0.972}	{0.304}	{0.653}
OtherXPost2	-0.067	0.005	-0.034	-0.032*	0.003	-0.013	0.009	0.001	-0.000
	(0.220)	(0.067)	(0.054)	(0.018)	(0.003)	(0.010)	(0.020)	(0.005)	(0.000)
			{0.647}	{0.453}	{0.647}	{0.453}	{1.000}	{1.000}	{1.000}
OtherXPost3	0.860***	-0.042	-0.064	-0.018	-0.010**	0.034*	0.042**	0.004	0.000
	(0.326)	(0.075)	(0.077)	(0.025)	(0.004)	(0.017)	(0.019)	(0.005)	(0.000)
			{0.315}	{0.315}	{0.060}	{0.083}	{0.099}	{0.766}	{0.766}
OtherXPost4	0.131	0.084	-0.010	-0.020***	-0.002	0.000	0.002	-0.008*	-0.000
	(0.085)	(0.057)	(0.029)	(0.006)	(0.003)	(0.003)	(0.015)	(0.004)	(0.000)
			{1.000}	{0.004}	{1.000}	{1.000}	{1.000}	{0.307}	{0.661}
Observations	1080	8107	1077	1077	986	1077	4663	8104	8104
Clusters	193	257	192	192	173	192	165	257	257
R^2	0.26	0.31	0.20	0.22	0.56	0.18	0.42	0.28	0.27
Control mean	0.08	0.05	0.23	0.07	0.03	0.01	0.41	0.05	0.00
Factory FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes: Table 2.E.20 shows treatment effects on HR and PR indices in column one and two, which refer to columns 11 and 12 in the summary Table 2.9, and individual index components (exit, turnover, absenteeism and promotion rates for HR index, and efficiency, alteration rates and reject rates for PR index). Regressions include time fixed effects to control for seasonal variations. We use weekly time dummies for the PR index and corresponding index components, and monthly time dummies for the HR index and corresponding index components. Factory fixed effects are also included to control for systematic differences across factories. The error term is clustered at the line level. False discovery rate-adjusted p-values, also known as q-values, for HR-related and PR-related outcome families were used to adjust for multiple outcome variables. They are shown in curly brackets. We use the 'sharpened q value' approach based on the Benjamini-Hochberg procedure (Benjamini et al., 2006). For simplicity, the regression table does not report coefficients on the $PostX$ variables for $X = (1, 4)$, and the baseline value of the dependent variables. Standard errors are shown in round parentheses. * p<0.1, ** p<0.05, *** p<0.01

TABLE 2.E.21. SPILLOVER EFFECTS: CONTROL FLOORS (COLUMNS 13 AND 14 IN SUMMARY TABLE 2.9)

	Index		**HR Variables**				**PR Variables**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	HR	**PR**	Exit	Turnover	Absenteeism	Promotion	Efficiency	Alterations	Rejects
TreatedXPost1	0.111	0.119	-0.053	-0.008	0.000	-0.001	0.012	-0.006	-0.000
	(0.095)	(0.104)	(0.040)	(0.006)	(0.004)	(0.002)	(0.011)	(0.006)	(0.000)
			{0.567}	{0.567}	{0.984}	{0.928}	{0.850}	{0.850}	{0.850}
TreatedXPost2	0.121	-0.023	-0.053	-0.033**	0.002	-0.005	0.015	0.003	0.000
	(0.202)	(0.060)	(0.046)	(0.014)	(0.003)	(0.010)	(0.017)	(0.004)	(0.000)
			{0.591}	{0.095}	{0.844}	{0.844}	{1.000}	{1.000}	{1.000}
TreatedXPost3	0.687**	-0.035	-0.048	-0.021	-0.003	0.029**	0.031*	0.002	0.000
	(0.297)	(0.072)	(0.068)	(0.023)	(0.004)	(0.013)	(0.017)	(0.005)	(0.000)
			{0.569}	{0.569}	{0.569}	{0.124}	{0.263}	{1.000}	{1.000}
TreatedXPost4	0.110	0.070	-0.017	-0.018***	-0.001	-0.000	-0.001	-0.006	-0.000
	(0.074)	(0.051)	(0.025)	(0.005)	(0.003)	(0.002)	(0.013)	(0.004)	(0.000)
			{1.000}	{0.005}	{1.000}	{1.000}	{0.534}	{0.534}	{0.534}
Observations	1080	8107	1077	1077	986	1077	4663	8104	8104
Clusters	193	257	192	192	173	192	165	257	257
R^2	0.25	0.31	0.19	0.21	0.56	0.18	0.42	0.28	0.27
Control mean	0.01	0.00	0.25	0.08	0.04	0.01	0.40	0.06	0.00
Factory FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes: Table 2.E.21 shows treatment effects on HR and PR indices in column one and two, and individual index components (exit, turnover, absenteeism and promotion rates for HR index, and efficiency, alteration rates and reject rates for PR index). Regression results are shown for factories of type "AB", which have control lines on different floors, to compare treatment floors to control floors. Columns one and two refer to columns 13 and 14 in the summary Table 2.9. Regressions include time fixed effects to control for seasonal variations. We use weekly time dummies for the PR index and corresponding index components, and monthly time dummies for the HR index and corresponding index components. Factory fixed effects are also included to control for systematic differences across factories. The error term is clustered at the line level. False discovery rate-adjusted p-values, also known as q-values, for HR-related and PR-related outcome families were used to adjust for multiple outcome variables. They are shown in curly brackets. We use the 'sharpened q value' approach based on the Benjamini-Hochberg procedure (Benjamini et al., 2006). For simplicity, the regression table does not report coefficients on the $PostX$ variables for $X = (1, 4)$, and the baseline value of the dependent variables. Standard errors are in round brackets. * p<0.1, ** p<0.05, *** p<0.01

TABLE 2.E.22. SPILLOVER EFFECTS: ALL FLOORS (COLUMNS 15 AND 16 IN SUMMARY TABLE 2.9)

	Index		**HR Variables**				**PR Variables**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	HR	**PR**	Exit	Turnover	Absenteeism	Promotion	Efficiency	Alterations	Rejects
TreatedXPost1	0.045	0.044	-0.038	-0.010*	0.002	-0.003	0.021*	-0.001	-0.000
	(0.091)	(0.093)	(0.036)	(0.006)	(0.004)	(0.002)	(0.011)	(0.005)	(0.000)
			{0.371}	{0.371}	{0.398}	{0.371}	{0.191}	{1.000}	{1.000}
TreatedXPost2	0.272	0.004	-0.081*	-0.033**	0.002	-0.000	0.019	0.002	0.000
	(0.185)	(0.055)	(0.044)	(0.013)	(0.003)	(0.009)	(0.016)	(0.004)	(0.000)
			{0.109}	{0.046}	{0.578}	{0.931}	{1.000}	{1.000}	{1.000}
TreatedXPost3	0.814***	0.016	-0.046	-0.051*	-0.002	0.027**	0.042**	-0.000	0.000
	(0.301)	(0.067)	(0.064)	(0.028)	(0.004)	(0.013)	(0.017)	(0.004)	(0.000)
			{0.392}	{0.156}	{0.392}	{0.156}	{0.039}	{1.000}	{1.000}
TreatedXPost4	0.117	0.052	-0.025	-0.016***	0.001	-0.001	-0.005	-0.006	-0.000
	(0.073)	(0.049)	(0.024)	(0.005)	(0.003)	(0.002)	(0.013)	(0.004)	(0.000)
			{0.786}	{0.005}	{1.000}	{1.000}	{0.849}	{0.587}	{0.800}
Observations	1245	9689	1245	1245	1154	1245	5579	9684	9684
Clusters	224	306	224	224	205	224	198	306	306
R^2	0.21	0.29	0.17	0.19	0.52	0.17	0.39	0.26	0.28
Control mean	0.02	-0.00	0.24	0.07	0.03	0.01	0.39	0.05	0.00
Factory FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes: Table 2.E.22 shows treatment effects on HR and PR indices in column one and two, and individual index components (exit, turnover, absenteeism and promotion rates for HR index, and efficiency, alteration rates and reject rates for PR index). Regression results are shown for factories of type "AB", which have control lines on different floors, to compare treatment floors to all other floors in the factory. Columns one and two refer to columns 15 and 16 in the summary Table 2.9. Regressions include time fixed effects to control for seasonal variations. We use weekly time dummies for the PR index and corresponding index components, and monthly time dummies for the HR index and corresponding index components. Factory fixed effects are also included to control for systematic differences across factories. The error term is clustered at the line level. False discovery rate-adjusted p-values, also known as q-values, for HR-related and PR-related outcome families were used to adjust for multiple outcome variables. They are shown in curly brackets. We use the 'sharpened q value' approach based on the Benjamini-Hochberg procedure (Benjamini et al., 2006). For simplicity, the regression table does not report coefficients on the $PostX$ variables for $X = (1, 4)$, and the baseline value of the dependent variables. Standard errors are in round brackets. * p<0.1, ** p<0.05, *** p<0.01

TABLE 2.E.23. ROBUSTNESS CHECK: FACTORIES WITH HIGH MATCH RATES BASED ON FLOOR INFORMATION

	All C-lines (ABC)		All C-lines		Diff C-lines		Adj-lines (ABC)		Adj-lines		Adj+Other		C-floor		All floors	
	(1) HR	(2) PR	(3) HR	(4) PR	(5) HR	(6) PR	(7) HR	(8) PR	(9) HR	(10) PR	(11) HR	(12) PR	(13) HR	(14) PR	(15) HR	(16) PR
TreatedXPost1	0.39	0.014	0.31	0.013	0.41	0.055	0.41	0.0052	0.42	0.054	0.32**	0.053	0.23**	0.14	0.12	0.045
	(0.26)	(0.20)	(0.26)	(0.30)	(0.27)	(0.31)	(0.25)	(0.20)	(0.25)	(0.30)	(0.16)	(0.28)	(0.11)	(0.14)	(0.11)	(0.12)
TreatedXPost2	0.46*	0.0026	0.55	0.0083	0.63	-0.048	0.49**	0.0020	0.64	-0.043	0.53	-0.12	0.18	-0.032	0.36	0.013
	(0.26)	(0.16)	(0.52)	(0.15)	(0.53)	(0.15)	(0.24)	(0.15)	(0.50)	(0.15)	(0.42)	(0.13)	(0.24)	(0.060)	(0.22)	(0.053)
TreatedXPost3	0.26	0.0059	0.37	0.028	0.50	-0.0011	0.29	-0.0087	0.51	-0.0047	0.76*	-0.079	0.80***	-0.068	0.88***	0.0012
	(0.34)	(0.17)	(0.39)	(0.19)	(0.41)	(0.20)	(0.35)	(0.16)	(0.43)	(0.19)	(0.41)	(0.18)	(0.28)	(0.084)	(0.30)	(0.076)
TreatedXPost4	-0.11	0.17	-0.23	0.15	-0.20	0.12	-0.094	0.17	-0.18	0.13	-0.12	0.11	0.089	0.052	0.11	0.047
	(0.19)	(0.12)	(0.28)	(0.14)	(0.29)	(0.14)	(0.18)	(0.12)	(0.28)	(0.14)	(0.24)	(0.12)	(0.081)	(0.061)	(0.083)	(0.055)
AdjacentXPost1							0.092	0.031	0.096	0.040	-0.022	0.059				
							(0.21)	(0.17)	(0.26)	(0.25)	(0.17)	(0.21)				
AdjacentXPost2							0.047	0.12	0.52	0.026	0.41	-0.069				
							(0.27)	(0.11)	(0.51)	(0.11)	(0.43)	(0.083)				
AdjacentXPost3							0.057	-0.060	0.30	-0.0013	0.49	-0.065				
							(0.30)	(0.14)	(0.37)	(0.17)	(0.32)	(0.15)				
AdjacentXPost4							0.049	0.13	0.14	0.066	0.18	0.070				
							(0.11)	(0.11)	(0.15)	(0.13)	(0.11)	(0.10)				
OtherXPost1											0.28**	0.23				
											(0.12)	(0.16)				
OtherXPost2											-0.031	-0.0045				
											(0.27)	(0.067)				
OtherXPost3											1.06***	-0.086				
											(0.29)	(0.088)				
OtherXPost4											0.12	0.027				
											(0.091)	(0.065)				
Observations	255	1697	182	1368	170	1275	408	2789	273	2144	969	6649	969	6649	1134	8231
R^2	0.44	0.35	0.43	0.39	0.40	0.40	0.40	0.33	0.40	0.39	0.34	0.36	0.34	0.36	0.28	0.33
Line clusters	45	56	32	43	30	40	72	92	48	67	174	208	174	208	205	257
Control Mean	-0.067	-0.069	-0.052	-0.0050	-0.019	-0.0044	0.016	-0.077	0.12	-0.013	0.042	0.016	-0.033	-0.033	-0.011	-0.0078
Factory Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Time Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes: Regressions exclude five factories with a match rate of less than 75% of floor information between survey and salary data. All indices are based on Kling et al. (2007). Columns 1 to 6 show regression results of treatment lines (T) compared to control lines, out of which columns 3 to 6 exclude factories with control lines only on the same floor as treatment lines. Columns 7 to 10 compare treatment lines (T) and adjacent lines (A) to control lines, out of which columns 9 and 10 exclude factories with control lines only on the same floor as treatment lines and only keep control lines on control floors. Columns 11 and 12 compare treatment lines (T), adjacent lines (A) and and other lines on the treatment floor (O) to all lines on the control floor. Columns 13 and 14 compare treatment floors to the control floors, and columns 15 and 16 against all lines on all other floors in the factory. Standard errors are shown in round parentheses. * p<0.1, ** p<0.05, *** p<0.01

TABLE 2.E.24. ROBUSTNESS CHECK: OVERALL EFFECTS, ALL CONTROL LINES (ABC) (COLUMNS 1 AND 2 IN SUMMARY TABLE 2.E.23)

	Index		**HR Variables**				**PR Variables**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	HR	**PR**	Exit	Turnover	Absenteeism	Promotion	Efficiency	Alterations	Rejects
TreatedXPost1	0.390	0.014	-0.081	-0.018*	-0.002	0.010	0.045*	0.002	0.000
	(0.257)	(0.204)	(0.063)	(0.009)	(0.008)	(0.016)	(0.025)	(0.011)	(0.001)
			{0.453}	{0.296}	{0.712}	{0.712}	{0.301}	{1.000}	{1.000}
TreatedXPost2	0.456*	0.003	-0.162**	-0.029	0.001	0.002	0.018	0.004	0.000
	(0.255)	(0.156)	(0.067)	(0.019)	(0.006)	(0.019)	(0.034)	(0.010)	(0.001)
			{0.088}	{0.256}	{0.854}	{0.854}	{1.000}	{1.000}	{1.000}
TreatedXPost3	0.263	0.006	-0.045	-0.015	-0.000	0.008	0.002	0.003	-0.000
	(0.338)	(0.167)	(0.171)	(0.014)	(0.008)	(0.013)	(0.045)	(0.010)	(0.000)
			{1.000}	{1.000}	{1.000}	{1.000}	{1.000}	{1.000}	{1.000}
TreatedXPost4	-0.112	0.167	0.028	0.004	0.002	-0.003	0.018	-0.003	-0.001**
	(0.193)	(0.124)	(0.041)	(0.008)	(0.006)	(0.006)	(0.030)	(0.009)	(0.000)
			{1.000}	{1.000}	{1.000}	{1.000}	{0.938}	{0.938}	{0.051}
Observations	255	1697	255	255	228	255	747	1696	1696
Clusters	45	56	45	45	40	45	32	56	56
R^2	0.44	0.35	0.31	0.39	0.53	0.25	0.56	0.30	0.34
Control mean	-0.07	-0.07	0.22	0.06	0.04	0.01	0.40	0.06	0.00
Factory FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes: Table 2.E.24 shows treatment effects on HR and PR indices in column one and two, which refer to columns 1 and 2 in the summary Table 2.E.23, and individual survey components (exit, turnover, absenteeism and promotion rates for HR index, and efficiency and alteration rates for PR index). Regressions include time fixed effects to control for seasonal variations. We use weekly time dummies for the PR index and corresponding index components, and monthly time dummies for the HR index and corresponding index components. Factory fixed effects are also included to control for systematic differences across factories. The error term is clustered at the line level. False discovery rate-adjusted p-values, also known as q-values, for HR-related and PR-related outcome families were used to adjust for multiple outcome variables. They are shown in curly brackets. We use the 'sharpened q value' approach based on the Benjamini-Hochberg procedure (Benjamini et al., 2006). For simplicity, the regression table does not report coefficients on the $PostX$ variables for $X = (1, 4)$, and the baseline value of the dependent variables. Regressions exclude five factories with a match rate of less than 75% of floor information between survey and salary data. Standard errors are shown in round parentheses. * p<0.1, ** p<0.05, *** p<0.01

TABLE 2.E.25. ROBUSTNESS CHECK: FACTORIES WITH HIGH MATCH RATES BASED ON LINE INFORMATION

	All C-lines (ABC)		All C-lines		Diff C-lines		Adj-lines (ABC)		Adj-lines		Adj+Other		C-floor		All floors	
	(1) HR	(2) PR	(3) HR	(4) PR	(5) HR	(6) PR	(7) HR	(8) PR	(9) HR	(10) PR	(11) HR	(12) PR	(13) HR	(14) PR	(15) HR	(16) PR
TreatedXPost1	0.25	0.098	0.18	0.038	0.29	0.12	0.28	0.100	0.28	0.13	0.24	0.10	0.26**	0.18	0.20*	0.12
	(0.22)	(0.41)	(0.23)	(0.49)	(0.25)	(0.49)	(0.21)	(0.41)	(0.22)	(0.49)	(0.15)	(0.43)	(0.10)	(0.17)	(0.10)	(0.16)
TreatedXPost2	0.43	0.098	0.53	0.11	0.61	0.096	0.47	0.10	0.59	0.10	0.47	-0.0055	0.22	-0.010	0.36	-0.020
	(0.43)	(0.17)	(0.53)	(0.18)	(0.55)	(0.19)	(0.40)	(0.17)	(0.52)	(0.18)	(0.44)	(0.16)	(0.24)	(0.065)	(0.22)	(0.061)
TreatedXPost3	-0.082	0.085	0.16	0.087	0.32	0.061	0.0094	0.089	0.29	0.068	0.56	-0.018	0.79***	-0.088	0.82***	-0.063
	(0.42)	(0.23)	(0.39)	(0.25)	(0.43)	(0.25)	(0.39)	(0.22)	(0.43)	(0.25)	(0.37)	(0.24)	(0.28)	(0.093)	(0.28)	(0.088)
TreatedXPost4	-0.18	0.11	-0.25	0.14	-0.22	0.11	-0.14	0.11	-0.22	0.10	-0.17	0.16	0.069	0.065	0.053	0.081
	(0.21)	(0.19)	(0.29)	(0.21)	(0.29)	(0.21)	(0.21)	(0.19)	(0.29)	(0.21)	(0.26)	(0.18)	(0.082)	(0.074)	(0.085)	(0.071)
AdjacentXPost1							0.044	0.067	0.12	0.039	0.072	0.023				
							(0.18)	(0.31)	(0.23)	(0.37)	(0.17)	(0.30)				
AdjacentXPost2							0.21	0.096	0.74	0.083	0.60	-0.017				
							(0.41)	(0.12)	(0.48)	(0.13)	(0.41)	(0.095)				
AdjacentXPost3							-0.65	-0.042	0.063	0.016	0.34	-0.055				
							(0.58)	(0.20)	(0.32)	(0.21)	(0.28)	(0.19)				
AdjacentXPost4							0.094	0.11	0.11	0.13	0.12	0.19				
							(0.13)	(0.16)	(0.16)	(0.18)	(0.11)	(0.14)				
OtherXPost1											0.35***	0.26				
											(0.12)	(0.20)				
OtherXPost2											0.0068	-0.0090				
											(0.27)	(0.071)				
OtherXPost3											1.05***	-0.12				
											(0.29)	(0.094)				
OtherXPost4											0.11	-0.0018				
											(0.092)	(0.078)				
Observations	196	1028	155	857	143	797	312	1626	224	1247	807	4701	807	4701	918	5364
R^2	0.40	0.38	0.45	0.40	0.43	0.42	0.38	0.36	0.43	0.42	0.37	0.37	0.36	0.37	0.32	0.35
Line clusters	34	34	27	27	25	25	54	54	39	39	144	147	144	147	165	168
Control Mean	0.013	-0.0076	-0.011	0.010	0.017	-0.021	0.082	-0.0037	0.14	-0.020	0.061	0.0031	-0.021	-0.045	0.019	-0.013
Factory Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Time Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes: Regressions exclude 13 factories with a match rate of less than 60% of line information between survey and salary data. All indices are based on Kling et al. (2007). Standard errors are shown in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Columns 1 to 6 show regression results of treatment lines (T) compared to control lines, out of which columns 3 to 6 exclude factories with control lines only on the same floor as treatment lines. Columns 7 to 10 compare treatment lines (T) and adjacent lines (A) to control lines, out of which columns 9 and 10 exclude factories with control lines only on the same floor as treatment lines and only keep control lines on control floors. Columns 11 and 12 compare treatment lines (T), adjacent lines (A) and and other lines on the treatment floor (O) to all lines on the control floor. Columns 13 and 14 compare treatment floors to the control floors, and columns 15 and 16 against all lines on all other floors in the factory. Standard errors are shown in round parentheses. * p<0.1, ** p<0.05, *** p<0.01

TABLE 2.E.26. ROBUSTNESS CHECKS: OVERALL EFFECTS, ALL CONTROL LINES (ABC) (COLUMNS 1 AND 2 IN SUMMARY TABLE 2.E.25)

	Index		**HR Variables**				**PR Variables**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	HR	**PR**	Exit	Turnover	Absenteeism	Promotion	Efficiency	Alterations	Rejects
TreatedXPost1	0.245	0.098	-0.057	-0.013	0.003	0.004	0.043	-0.006	0.001
	(0.215)	(0.414)	(0.075)	(0.010)	(0.009)	(0.005)	(0.027)	(0.051)	(0.002)
			{1.000}	{1.000}	{1.000}	{1.000}	{0.536}	{1.000}	{1.000}
TreatedXPost2	0.431	0.098	-0.211**	-0.034	0.008	0.000	0.050	0.009	-0.000
	(0.433)	(0.170)	(0.093)	(0.025)	(0.008)	(0.025)	(0.038)	(0.025)	(0.001)
			{0.142}	{0.362}	{0.494}	{0.788}	{1.000}	{1.000}	{1.000}
TreatedXPost3	-0.082	0.085	0.139	-0.015	-0.003	0.002	0.012	-0.005	-0.001
	(0.418)	(0.225)	(0.210)	(0.015)	(0.009)	(0.011)	(0.052)	(0.031)	(0.000)
			{1.000}	{1.000}	{1.000}	{1.000}	{1.000}	{1.000}	{1.000}
TreatedXPost4	-0.179	0.108	0.013	0.012	0.007	-0.004	0.006	0.000	-0.001*
	(0.214)	(0.186)	(0.049)	(0.011)	(0.007)	(0.005)	(0.023)	(0.027)	(0.001)
			{1.000}	{1.000}	{1.000}	{1.000}	{1.000}	{1.000}	{0.232}
Observations	196	1028	196	196	173	196	527	1027	1027
Clusters	34	34	34	34	30	34	24	34	34
R^2	0.40	0.38	0.33	0.39	0.55	0.23	0.56	0.38	0.27
Control mean	0.01	-0.01	0.26	0.06	0.04	0.01	0.37	0.08	0.00
Factory FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes: Table 2.E.26 shows treatment effects on HR and PR indices in column one and two, which refer to columns 1 and 2 in the summary Table 2.E.25, and individual survey components (exit, turnover, absenteeism and promotion rates for HR index, and efficiency and alteration rates for PR index). Regressions include time fixed effects to control for seasonal variations. We use weekly time dummies for the PR index and corresponding index components, and monthly time dummies for the HR index and corresponding index components. Factory fixed effects are also included to control for systematic differences across factories. The error term is clustered at the line level. False discovery rate-adjusted p-values, also known as q-values, for HR-related and PR-related outcome families were used to adjust for multiple outcome variables. They are shown in curly brackets. We use the 'sharpened q value' approach based on the Benjamini-Hochberg procedure (Benjamini et al., 2006). For simplicity, the regression table does not report coefficients on the $PostX$ variables for $X = (1, 4)$, and the baseline value of the dependent variables. Regressions exclude 13 factories with a match rate of less than 60% of line information between survey and salary data. Standard errors are shown in round parentheses. * p<0.1, ** p<0.05, *** p<0.01

CHAPTER 3

Does the Devil wear Zara? - The Impact of Fast Fashion and Low-level Autonomy in the Bangladeshi Garment Industry

with **Ferrán Vega and Christopher Woodruff**

Abstract This paper analyses detailed line-level data from a Bangladeshi garment factory to examine the role of low-level autonomy in the process of adaptation to a negative productivity shock. We begin by qualitatively describing (i) the productivity shock caused by switching to a 'faster fashion' production model, and (ii) an impressive, but uneven, recovery despite unchanged conditions. We then examine those trends. First, our data suggests that the recovery was triggered by learning processes, for example by building up 'working knowledge' at the level of the individual item being produced. Second, our data suggests that performance differences across units can be traced back to varying degrees of autonomy exercised by production managers in deviating from pre-determined plans. After comparing the efficiency costs in the production data with those embedded into the plans, we find that the change in the factory's production model undermined the planning departments' ability to generate adequate production plans. We argue that re-organising production beyond plans was fundamental to keep up with a high-paced industry.

3.1 Introduction

Many lower-income countries are characterised by both low levels of trust and highly volatile business environments. For firms operating in these environments, this creates a tension between a desire to centralise decision-making in response to the low levels of trust (Bloom et al., 2012) and the need for flexibility to deal with frequent shocks that arise from the more volatile business environment. Centralised decision-making arises because incentives of lower-level managers are often not aligned with those of the firm and owners and top managers do not trust lower-level managers to make decisions in the owners' interest. But firms benefit from delegating more decision-making authority to lower-level managers when information that allows better adaptation to shocks is dispersed and held at lower levels.

We examine this tension using data from a very large export-oriented Bangladeshi garment factory. Bangladesh is a low-trust country,[1] and decision-making in factories of the type we study is highly centralised. We examine a period of time (2015-2020) when the firm faced a rapid decline in the average size of orders, a shortening of lead times and a concurrent increase in the number of orders - a move from "fast fashion" to "faster fashion". We compare adaptation as revealed by monthly production plans and by the deviations from those plans. We take the production plans to represent centralisation: high-level managers can view and assign production for an entire month across the factory's four groups containing 13 production units and 118 production lines. But "there is many a slip 'twixt the cup and the lip": shocks from late delivery of fabric, machine outages, and worker absences mean that sticking to the plans will not be optimal. The observed deviation from those plans represents the "real authority" (Aghion and Tirole, 1997) of lower level managers. This may be either beneficial to the firm - managers using detailed knowledge about conditions in their unit or on their line to adapt the plan to changing circumstances - or detrimental, as the lack of incentive alignment leads lower-level managers to use shocks in their units to shift rents down the hierarchy.

[1] Bangladesh is a low-trust country (WVS, 2020) When asked "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?", 87% of respondents report the need to be "very careful".

Beginning in mid-2016, the factory's well-known brand-name customers sharply reduced the average size of orders and decreased the time available to complete them. These changes were the consequence of industry-wide efforts to speed up supply chains and decrease the amount of inventory held at any point in time, as documented in the trade journals in relation to giants in the industry such as H&M (Kondracki, 2018). This move to "faster fashion" was driven by competition with the model of Inditex (whose best known brand Zara located production much closer to market), while retaining the cost advantage of producing in low-wage countries like Bangladesh. In the factory that is our main focus, the change coincided with a 33 percent decrease in productivity relative to its 2015 level. Though the change in order pattern endured, the factory's productivity recovered so that by 2019, efficiency was within 12 percent of the 2015 level. However, the recovery was very uneven across the units and production lines in the factory, with a significant widening of the gap between the most- and least productive units. We link this pattern to differences in the degree of autonomy displayed by the factory's production units in deviating from pre-determined plans and provide evidence that allowing workers to exercise "real authority" within well-defined organisational entities can be productivity-enhancing.

Our data are observational, but given the level of detail and the abrupt nature of the changes external to the factory, we believe they are informative about the adaptation process. Our approach is two-fold. First, we measure the effects of external and internal production determinants on productivity- and quality-related outcomes, i.e. efficiency, alteration rates and rejection rates, comparing the magnitude of these effects to an industry benchmark. The external factors are those which the factory cannot directly influence, primarily the shock related to the timing, quantity, and complexity of the orders from buyers. The internal factors relate to how the factory distributes the workload from product orders across production lines. We show that efficiency and quality vary systematically with characteristics of the style and production processes. Efficiency is lower, and quality defects higher, for more complex styles and when the factory produces multiple styles on the same production line on a given day. There are also strong style-specific learning effects (Menzel, 2020). For learning at the style level, we differentiate between

'cumulative knowledge', which increases with each additional day a style is produced on the line or in the unit within and across production spells and 'working knowledge', which increases for each style within a given production spell. We find that the factory has high productivity gains from building 'working knowledge', whereas effects from building 'cumulative knowledge' at the line or unit level are negligible.

Second, we measure separately the efficiency costs from internal factors related to the allocation of styles to production lines embedded in the monthly production plans and in the actual production data.[2] These are costs associated with (i) producing multiple styles on a given line on a given day, (ii) producing less similar styles on a given day, and (iii) producing less similar styles on a line from one day to the next.[3] Together, the two measures provide insight into communication between the planning and production departments, and the "real authority" (Aghion and Tirole, 1997) exercised by the production managers in each unit. As in other low-trust countries (Bloom et al., 2013), garment factories in Bangladesh have highly centralised decision-making structures, with decisions being made almost exclusively in headquarters located outside the factory.[4] One way for higher-level managers to maintain control of daily production from headquarters is to produce detailed plans.[5] However, plant managers often have an informational advantage and, in spite of agency issues, it may be beneficial to allow these lower-level managers the autonomy to use it. In turn, using this local information can result in deviation from plans and undermine "formal authority", a dynamic that is at the core of our analysis.[6] A qualitative comparison of the deviations from the plans by production unit reveals that higher-performing units deviate *more*

[2]We have production plans for each month, each covering a time horizon of one month.

[3]Managers have control over these three factors on a daily basis and regressions indicate that they have effects on efficiency.

[4]In a small sample of large factories with headquarters located outside the factory, 100 percent of plant managers said they needed approval from headquarters for expenditures as small as $1200 (Surveys conducted in the sample described in Uckat and Woodruff (2020)).

[5]Exercising control through planning documents is easier, because there is a single file for the entire month for the whole factory, which managers can look at at a point in time. Production reports, however, are much more dispersed across production units.

[6]The need to deviate from production plans may arise either because of last minute changes in the orders that could not be reflected in the planning data, from failure to keep to plan in departments other than production (e.g., the fabric production department), or failure to keep to plan in the production units. The scale of the latter two issues is likely to grow as the factory grapples with the adaptation to a new production model.

from plans than lower-performing units, providing more flexibility in production. Further, we find that those units seem to cooperate more with other units within their group, as lines within more productive organisational groups re-shuffle orders to increase the average share of styles that are produced on the exact day they were planned (across all lines in the group). In turn, this suggests that the exercise of 'real authority' by lower level managers is beneficial to the factory in the process of adaptation, in spite of any agency concerns. A quantitative comparison of the efficiency costs from the three factors mentioned above across planning and production reports confirms this finding. The data show that the production department first attempted to adhere to the new planning strategy after the changes in order flow, but increasingly deviated from plans as they found themselves unable to adjust. Over time, the planning department appears to learn from the production department and update the planning strategy, after which the gap between production and planning closes. These patterns are more pronounced in the higher-performing units than in lower-performing units, magnifying the performance differences within firms.

This paper contributes to three strands of the literature. First, we contribute to the non-experimental literature on determinants of productivity differences, how specific productivity gains accrue at the plant level, and how factories adapt to external shocks that negatively affect productivity. Performance differences have been documented among seemingly similar firms and within firms across a wide range of industries (Bartelsman and Doms, 2000; Hsieh and Klenow, 2009; Syverson, 2004, 2011). However, despite the relevance of this research, the question of what causes these persistent differences remains largely open due to difficulties in obtaining detailed firm-level data over longer periods of time. Here, our main contribution is the use of rich line-level data to provide a detailed decomposition of productivity into internal and external determinants and estimate their effects on productivity- and quality-related outcomes.

Second, we contribute to the literature on learning and adaptation within and across firms (Thornton and Thompson, 2001; Hanna et al., 2014; Hendel and Spiegel, 2014; Menzel, 2020; Atkin et al., 2017). To shed light on productivity growth, the literature has analysed learning within and across production plants

by quantifying the evolution of productivity and its determinants (Hendel and Spiegel, 2014; Levitt et al., 2013).[7] We contribute to this literature by studying two types of learning, working knowledge and cumulative knowledge within firms, thereby providing a way to separate the productivity effects of flow and stock measures of industry-dependent knowledge.

Third, while our contribution is empirical, our work speaks to the importance of delegation when there is important information that resides at lower levels, whether that delegation is explicit (Bloom et al., 2012) or implicit (Aghion and Tirole, 1997). In our factory, higher-level management can exercise control through the planning documents more easily than through production. Yet allowing production managers the autonomy to deviate from the plan is important to minimize the costs from shocks.[8] Such deviations may also increase cross-department learning from production to planning, allowing for feedback showing that the plans can be improved. While Aghion and Tirole (1997) focus on agency issues, it is worth noting that the inefficiency in adaptation may be passive rather than active (Almunia et al., 2019; Bandiera et al., 2009), that is, arising from a lack of competence rather than agency.[9]

After a description of the data in Section 3.2, we qualitatively describe current trends in the main factory in Section 3.3. Section 3.4 explains the empirical strategy, whereas Section 3.5 shows the effects of production determinants on productivity- and quality-related outcomes, and their interpretation. Section 3.6 provides a detailed analysis on what we can learn from differences between planned and actual production. Section 3.7 concludes.

[7]For example, Hendel and Spiegel (2014) find that learning by experimentation explains slow and steady productivity growth in steel mills. Similarly, Levitt et al. (2013) investigate learning by doing effects in an auto producer's assembly plant, finding that learning does not stay with individual line workers, but rather becomes embodied in the physical or broader organisational capital of the plant.

[8]Costs may arise from any combination of external and internal pressures described above, i.e. outside the factory, inside the factory but outside the production unit, or inside the production unit.

[9]Our work also closely links to theoretical papers on productivity differences within firms and countries (Gibbons and Henderson, 2012; Garicano and Rossi-Hansberg, 2015; Garicano and Rayo, 2016). For example, we provide empirical evidence on phenomena described by Garicano and Rayo (2016) in their paper about why organisations fail. For example, we observe productivity differences across units because of the inability of the production department in lower-performing units to allow additional flexibility in production, which Garicano and Rayo (2016) would categorise as failures due to a lack of adaptation to change (bound by rigid routines and cultures) or by a lack of communication.

3.2 Data

3.2.1 Selection of sample factories

We rely primarily on data from a single large garment factory employing nearly 10,000 workers on more than 100 production lines. The factory exports all of its production to several of the largest and most demanding buyers sourcing in Bangladesh. We focus on this factory for three reasons. First, the factory faced a significant shock to their order flow between 2016 and 2018. The change in order flow required adaptation by the factory, and we aim to understand how and how well the factory navigated this process. Second, the factory maintained digital records at a granular level over a period covering two years before and up to two years after the change in order flows. This allows us to conduct a detailed analysis over a long time horizon. Third, in addition to the availability of high-quality data, the factory is particularly well-suited for our analysis because different units (production floors) produce comparable products while having relatively autonomous management teams. The factory produces light knit products, almost exclusively t-shirts, tops and pajamas, which are simple and comparable across production lines.

The factory experienced changes to its management structure prior to the period of our analysis, but maintained a relatively stable management hierarchy throughout the years covered by our data. Between 2009 and 2013, the factory was jointly managed by the Managing Directors of the different departments. In mid-2013, they appointed their first external CEO. Since that time, the turnover of CEOs has been rather high: The first CEO stayed only 18 months until the end of 2014; the second CEO stayed for two and a half years from 2015 until mid-2018; and the third CEO has been in office since 2018.

Although most of the analysis focuses on data from this single factory, we also use data from one other factory to provide a rough control for broader industry factors that might affect outcomes.[10] We refer to this factory as our "industry benchmark". It is of similar size, focuses on similar product categories, and also made

[10]Future versions of the paper will use a larger group of comparison factories.

its granular production data available to us. Both our main and comparable factory allow us to calculate efficiency for multiple styles on the same day separately, which enables us to carry out the analysis at the style-day level while assuring the accuracy of comparisons across the two factories.

3.2.2 Data sources, outcome variables and summary statistics

We use administrative records from the Production and Planning departments of the Sewing Section in the factories.[11] Both factories have dedicated Industrial Engineering ("IE") departments that manage most of the technical aspects of production including setting targets and determining the layout of the line. Production data includes daily line-level information on the specific styles produced, a measure of the complexity of those styles, the allocated and actual manpower in minutes, the number of pieces produced, and quality-related information such as the number of pieces checked for quality, and the number rejected or requiring alterations. The Planning Department produces line-level plans for a one-month window (though some planning reports cover as many as 60 days). We collected the last plan produced in each calendar month. These plans include information on the planned style and planned production quantity for each line for each day during the planning window. The planning data also tell us, for each style, the delivery date, the total delivery quantity, and the planned sewing start and completion date.

Table 3.1 shows an overview of key variables from the production data, including all those entering our empirical specification, for both the core factory and the comparison factory. Column 1 of this table shows averages of the variables across the full sample period from 2015 to 2020 for our main factory.

To analyse how and to what extent performance differences are driven by external and internal factors, we investigate effects on three productivity- and quality-related outcomes: (i) efficiency, defined in the industry-standard manner as the number of output minutes (standard minute value multiplied by the daily output) divided by the number of available minutes, (ii) alteration rates, defined as the percentage of pieces per day that require additional work to fix a defect, and (iii)

[11] We have only production data for the comparison factory.

TABLE 3.1. SUMMARY STATISTICS

	Main Factory				Comparison Factory			
	Full period	2015	2017	2019	Full period	2015	2017	2019
Outcomes								
Efficiency	0.45	0.51	0.34	0.45	0.49	0.49	0.46	0.47
Alteration rate	0.08	0.07	0.07	0.11	0.07	0.05	0.09	0.09
Rejection rate	0.006	0.004	0.007	0.006	0.006	0.007	0.005	0.004
Daily output	1020	1327	772	911	1409	1497	1239	1242
External determinants								
Style SMV	6.58	7.38	6.78	5.89	7.40	8.39	7.27	6.68
Internal determinants related to style transitions								
Number of styles produced each day by line	1.78	1.51	1.74	1.88	1.36	1.31	1.30	1.47
Percentage of multiple style days per month	0.38	0.28	0.37	0.41	0.19	0.16	0.15	0.25
SD of SMV (line/day level)	0.47	0.35	0.48	0.48	0.26	0.30	0.24	0.29
Proportional change in SMV day average	0.07	0.05	0.07	0.07	0.04	0.04	0.04	0.04
Internal determinants related to production knowledge								
Running days - Reset at 7	5.24	5.38	5.69	4.98	6.33	6.68	6.59	5.98
Days passed since style X was produced	1.32	1.11	1.29	1.53	1.63	1.40	1.59	2.07

rejection rates, the proportion of items with defects that cannot be repaired. Efficiency averages 45 percent in the main factory and 49 percent in the comparison factory, though efficiency may not be directly comparable across factories because of different standards used by industrial engineers. Our focus factory produces 1,020 pieces per line-day against 1,409 pieces in the comparison factory. Alteration rates average 8 percent in the main and 7 percent in the comparison factory, while reject rates are 0.6 percent in both factories. This indicates that, overall, differences between the two factories are small.

As external determinants of production, we consider factors which are imposed upon the factory by the industry's demand, such as the timing, quantity and complexity styles included in orders. Complexity is measured as the standard minute value (SMV). The SMV is estimated by industrial engineers prior to the start of production. It is the sum of the time, in minutes, required to perform each of the sewing operations needed to assemble one piece of the style. The engineers assume production under ideal conditions by a highly-skilled operator. Thus, the SMV provides a measure of the required labour input per piece, and a higher SMV indicates that either more steps, or more complex steps, are required to produce the garment. On average, the SMV is 6.58 minutes in the main factory and 7.40 minutes in the comparison factory.

We consider as internal determinants of production those factors which can be

changed by management on a daily basis. Factory managers have the discretion to allocate a given set of orders to production lines, a process that determines the production spells of each style on each line. For example, the factory can choose to produce a style on two different production lines for one week each or on one production line for two weeks. As efficiency is lower on the first days a style is produced on a line, the latter can be less costly in terms of the ramp-up costs. Efficiency is also affected by the similarity of consecutive styles produced on a given line. Transitioning to a style that is very different from the one previously produced on the line requires a greater number of changeovers in machines and workers. We measure the efficiency effect of the allocation of styles to lines with two sets of variables that track style transitions within and across days. We measure (i) the number of styles produced per line on a given line, (ii) the variation in style complexities (SMVs) per line on a given day, and (iii) the variation in style complexities per line from one day to the next.

Our focus factory often produces more than one style on a single line on a given day. In any factory, there will usually be two styles produced on days when the line is completing the production of one style and transitioning to another style. However, lines may also produce multiple styles when delivery dates are near and production of a given style needs to be rushed. Our focus factory produces on average 1.78 styles per day on a line (compared with 1.36 styles in the comparison factory), and produces more than one style per day almost 40 percent of the production days. Given that the average maximum production spell is around 13 days,[12] this evidently reflects more than simply style transitions. The efficiency cost of producing multiple styles on a given line-day depends on how different those styles are. A second style may be a minor variation on the first, adding a pocket on a t-shirt, for example. The efficiency cost will be greater if the two styles are very different from one another. We measure the difference between styles produced on the same line-day as the standard deviation of the style-specific SMVs per line-day. In our focus factory, this averages 0.47, which is 7 percent of the average SMV across the sample period. Again, this number is much lower in the industry benchmark.

[12]This is an average of the maximum production spell length across all styles, weighted by style occurrence.

Efficiency is also affected by the change in styles being produced from one day to the next. Here we measure the change as the average absolute difference in the average of style-specific SMVs from one day to the next. Again, this is larger in our focus factory (0.07) than in the comparison factory (0.04).

Finally, efficiency is typically lower on the first days a style is produced on a line, or in a factory (Menzel, 2020). We track the number of days each style is produced (running days) on a line. Where a line stops and then re-starts production of the same style we measure the length of the gap between production spells (days passed). Since operators' knowledge about how to produce a given style is plausibly not affected by going a few days without producing that style, we only reset the running days back to one if the gap in days between the day a line last produced the style and the day it is produced again is a week or more. The average across the full sample period in the main factory for running days and days passed is 5.24 days and 1.32 days, respectively. The running days average is higher in the comparison factory, indicating the faster style turnarounds in the focus factory.

Overall, these descriptive statistics suggest that the focus factory experiences a larger number of style transitions within and across days, and a lower number of consecutive production days on any given style.

3.3 Qualitative description of trends over time

In this section, we describe qualitatively two core features of the data, which build the motivation for our research and allow us to investigate drivers of variance in production outcomes. The first is the shock external to the factory that generated the need to adapt: the move to even "faster fashion" coming from changes in the nature of the demand from buyers. We show that a substantial drop in average order size was followed by sharp productivity losses in 2017. These were then followed by a substantial recovery in 2018, the recovery coming without any additional shocks to the factory's production model. Second, we show that the shock to order sizes comes with increasing heterogeneity across units within the same factory. In doing so, we focus on two production units (containing 20 production lines) that appear to have adapted rapidly to the changes and one unit (containing 10 production lines) where efficiency fell and never rebounded.

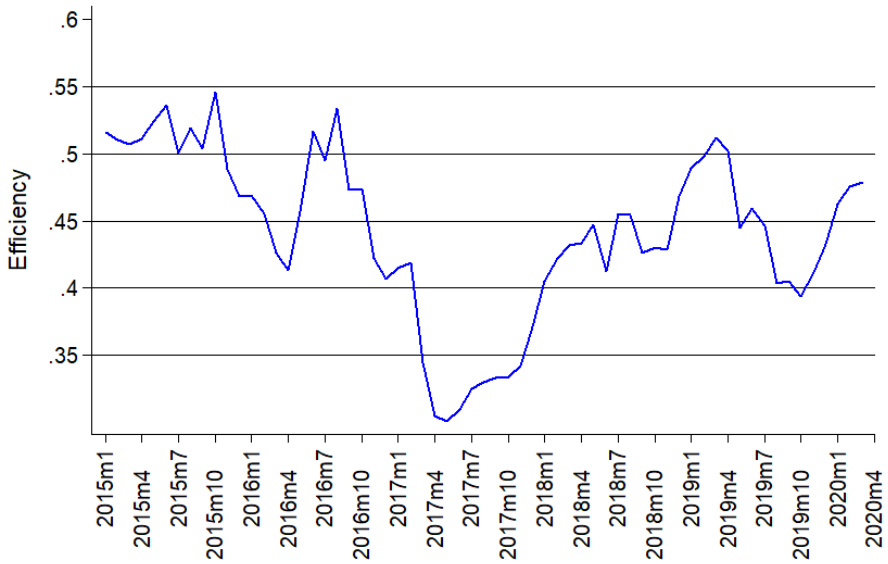


Figure 3.1: Efficiency over time

Notes: Table 3.1 shows the evolution of average efficiency in the study factory over time.

Figure 3.1 shows that the factory experienced large efficiency losses in 2017, with the efficiency measure falling from a yearly average of 51 percent in 2015 to 34 percent in 2017 (see summary statistics in Table 3.1), which is a fall of 17 percentage points, or 33 percent.

In early 2018, we observe a remarkable recovery of efficiency back to levels comparable to those preceding the plunge. The data show similar patterns for alteration rates and reject rates (see appendix graph 3.A.1). In contrast, productivity- and quality-related outcomes remained largely stable across time in the comparison factory (see Table 3.1). With the trend of "faster fashion" setting in stronger in our main factory, operators and managers faced shorter lead times, smaller order sizes, and a higher number of order. In fact, interviews with factory personnel confirmed that the second CEO, who was in office from 2015 until mid-2018, made a conscious management decision to change the factory's production model to adapt to these trends.

The upper panel of Figure 3.2 illustrates the sharp decrease in the factory's average delivery quantities, which reached their lowest point over the study period in 2017. The lower panel shows a simultaneous large increase in the total num-

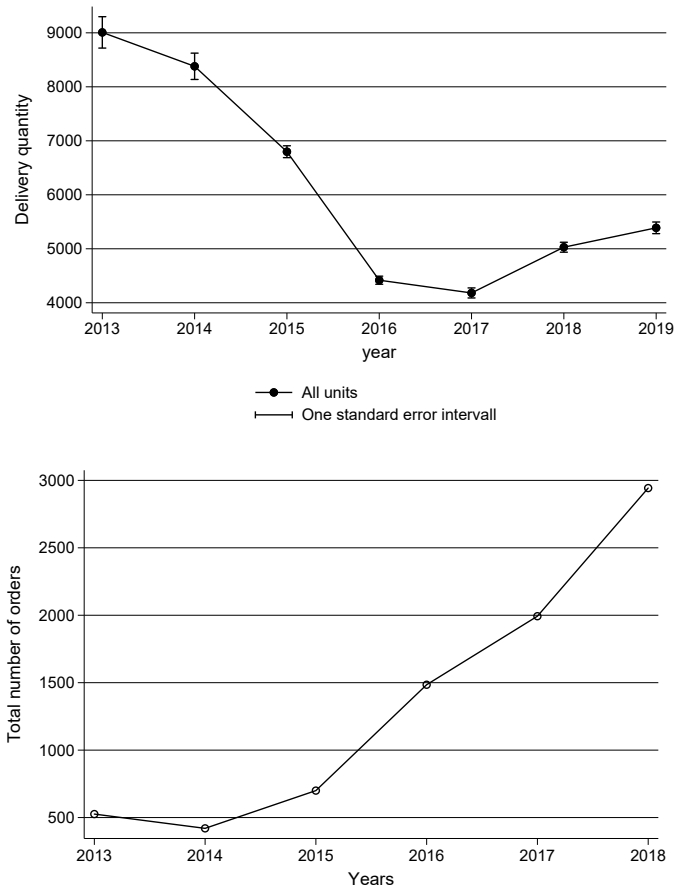


Figure 3.2: Imposed industry demand

Notes: The upper graph of Figure 3.2 shows the median and standard deviations of delivery quantity across years. Delivery quantity is calculated as the sum of daily planned quantities per style and delivery date. The lower graph shows the total number of orders across years in the whole factory. In 2019, 79 percent of the order IDs are missing, which implies that the total order size is only based on a small share of all orders in that year. Hence, we do not show the total number of orders in 2019.

ber of orders. At the same time, the factory seems to produce products in a much shorter time. In fact, we can observe a decrease in the actual sewing time, i.e. the number of days between the planned sewing start time and planned sewing completion time (see appendix graph 3.A.3), also consistent with the reduction in running days seen in Table 3.1. In contrast, we observe relatively small changes in the total delivery time, which we define as the actual sewing time plus the days until the product has to be delivered to the customer, i.e. the number of days between the planned sewing start and scheduled delivery date (see appendix graph 3.A.4. This pattern of decreasing actual sewing time and a relatively stable total

delivery time suggests that the factory takes more days for shipping into account, possibly due to increased pressure from buyers to deliver products on time. The trends in delivery quantities and the timing of the orders remain largely stable after 2017, which we take as evidence that the factory had finished the process of adapting its production model to the trend of "faster fashion".

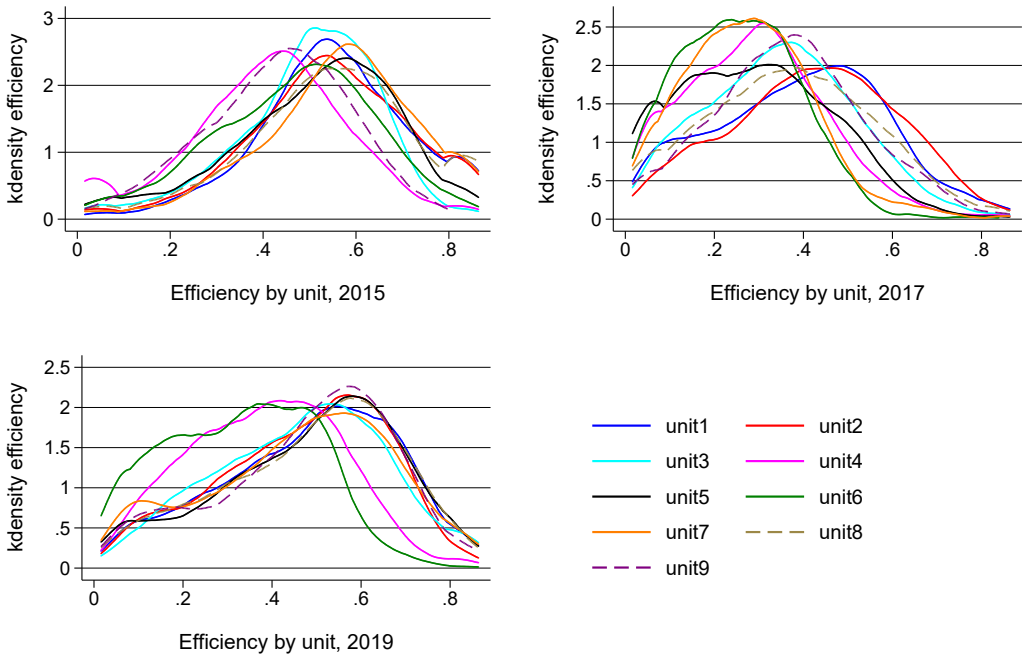


Figure 3.3: Efficiency dispersion (2015, 2017 and 2019)

Notes: Figure 3.3 shows the density of average efficiency across units in 2015, 2017 and 2019.

Zooming in at the unit level reveals increasing dispersion in efficiency and quality over time. The density plots of efficiency across units in Figure 3.3 clearly show much larger dispersion is in 2017 than in 2015. Dispersion decreases by 2019, but remains higher than in 2015. The increased dispersion from 2015 to 2019 comes from the fact that some units, such as Units 4 and 6, lag well behind in 2019, with large left tails indicating days with very low efficiency. In contrast, Units 1 and 2 perform consistently well - even in 2017. One possible explanation for these patterns is that the factory may assign different styles to each unit, and some units may produce styles that were affected more by the changes in order patterns than others. The data do show that, in 2015, some units produced different products, on

average, than other units. For example, Table 3.B.4 shows that the average SMV (complexity) of styles produced in 2015 in Unit 4 (9.35) was much higher than those produced in Units 1 or 2 (6.89 and 7.14, respectively). However, Table 3.3 shows that the complexity of styles produced in Unit 6 was quite similar to those produced in Units 1 and 2 both in 2015 and in 2017. More to the point, the drop in efficiency in Unit 6 relative to Units 1 and 2 holds even if we restrict the data to styles produced in all three units (see Figure 3.4). This suggests that allocation of styles across units does not explain the performance changes we observe in the data.

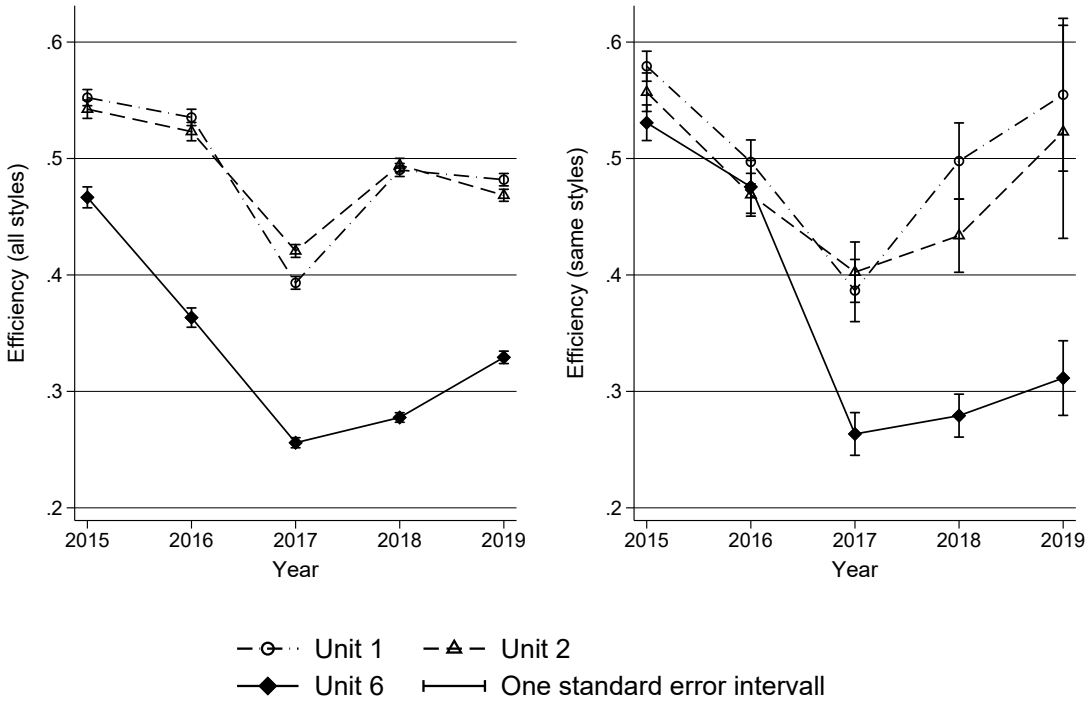


Figure 3.4: Efficiency over time for same styles and all styles

Notes: Figure 3.4 shows the mean and standard deviations of efficiency across years for Units 1, 2 and 6. The left graph includes all styles, whereas the right graph includes only 110 styles which are produced in all three units (at least once).

Regressing efficiency against a set of controls and style fixed effects reveals a similar pattern. The result in Table 3.2 (see extended version in Table 3.B.9) show that Units 1 and 2 outperform Unit 6 (set as the base category) throughout the entire time period. Columns 3, 5 and 7 compare the relative performance of the units

TABLE 3.2. EFFECTS ON EFFICIENCY USING ALL STYLES AND COMMON STYLES FOR SELECTED UNITS

	All styles (Full period)	Common styles (Full period)	All styles (2015)	Common styles (2015)	All styles (2017)	Common styles (2017)	All styles (2019)	Common styles (2019)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Unit=1	15.145***	7.683***	8.073***	5.537***	12.363***	9.267***	15.769***	2.368
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.793)
Unit=2	15.157***	6.891***	7.732***	5.519***	14.644***	7.553***	14.208***	12.020*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.085)
Unit 1=Unit 2	0.94	0.19	0.46	0.98	0.00	0.34	0.00	0.16
Line FE	No	No	No	No	No	No	No	No
Month/year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Style FE	No	Yes	No	Yes	No	Yes	No	Yes

Notes: This table is equivalent to Table 3.B.9 in the appendix, but we only keep the unit coefficients, where the base category is the low performing Unit 6. All of the regressions include controls for complexity . Odd columns use all styles and include unit, month/year, and buyer fixed effects. Even columns use only styles that are produced in all three units and also include style fixed effects. * p<0.1, ** p<0.05, *** p<0.01

in 2015, 2017 and 2019, respectively, with the full sample of styles. Columns 4, 6, and 8 show the same three years in a regression that includes style fixed effects using the subsample of styles produced in all three units. The gaps in performance are somewhat smaller in the regressions with the sample limited to common styles, but the pattern is the same. The differences in performance not only persist, but increase during 2017.[13] This observation is also illustrated in Table 3.3, which highlights percentage differences between Unit 6 and Units 1 and 2 of more than 20 percent in red. Differences in 2015 are smaller than 20 percent for all production-related variables, irrespective of whether we restrict to their common styles or not. In 2017, however, efficiency and quality gaps are much larger and the subsequent recovery in Units 1 and 2 fails to materialize in Unit 6. These patterns hold despite other production-related variables, such as SMV, remaining similar in all three units. Furthermore, the trends of the delivery quantities and the order sizes are also similar (see Figure 3.A.2).

Overall, the factory managed to recover after massive efficiency losses in 2017 despite additional pressure received from buyers or changes to their production model, with the caveat that heterogeneity across units remains. Particularly noteworthy are differences in productivity- and quality-related outcomes between the

[13]Formal tests show that Units 1 and 2 do not differ significantly in performance for the full period or prior to the shock in 2015. This provides additional reassurance for the validity of combining Units 1 and 2 for comparison to Unit 6.

TABLE 3.3. OVERVIEW OF VARIABLES RELATED TO PRODUCTION

		All styles				Same styles			
	Year	Units 1 and 2	Unit 6	Delta in %	N	Units 1 and 2	Unit 6	Delta in %	N
Efficiency	2015	55%	47%	-15%	5652	57%	53%	-7%	1418
	2017	41%	26%	-37%	12981	39%	26%	-33%	494
	2019	48%	33%	-31%	14182	54%	31%	-42%	111
Alteration rates	2015	7%	8%	4%	7843	7%	7%	7%	1926
	2017	6%	7%	27%	12154	6%	7%	12%	490
	2019	8%	18%	112%	13469	9%	15%	59%	110
Reject rates	2015	0%	1%	41%	7358	0%	1%	50%	1792
	2017	1%	1%	-33%	6266	1%	1%	-1%	282
	2019	1%	1%	33%	12444	0%	1%	105%	85
Running days (capped at 7)	2015	5,7	4,8	-15%	9377	5,1	5,2	2%	2362
	2017	6,0	5,1	-16%	13117	5,4	5,5	3%	496
	2019	4,7	4,8	2%	14225	5,4	5,0	-8%	111
Days passed	2015	1,2	1,1	-7%	9377	1,1	1,1	-8%	2362
	2017	1,3	1,4	3%	13117	1,2	1,1	-9%	496
	2019	1,7	1,5	-10%	14225	2,2	6,8	206%	111
Style SMV	2015	7,0	7,4	6%	9377	6,4	6,5	2%	2362
	2017	6,1	6,2	2%	13117	5,4	5,9	9%	496
	2019	5,1	6,6	30%	14225	6,3	5,1	-20%	111
SD SMV (line level)	2015	0,4	0,4	2%	9377	0,4	0,3	-20%	2362
	2017	0,4	0,4	-1%	13117	0,5	0,3	-40%	496
	2019	0,4	0,5	31%	14225	0,2	0,4	110%	111
Number of styles produced each day by line	2015	1,6	1,5	-6%	9377	1,6	1,5	-8%	2362
	2017	1,9	2,0	2%	13117	2,0	1,7	-14%	496
	2019	2,2	2,0	-9%	14225	2,0	1,6	-16%	111
Proportional change in SMV day average	2015	0,1	0,1	1%	13028	0,1	0,0	-27%	2346
	2017	0,1	0,1	5%	13117	0,1	0,1	-19%	496
	2019	0,1	0,1	8%	14225	0,0	0,1	39%	111

Notes: Table 3.3 displays yearly averages of key variables from the production data. It shows averages across Units 1 and 2, as well as Unit 6 for the years 2015, 2017 and 2019. The column "Delta in %" shows the delta between Units 1 and 2 and Unit 6 in percentage points for the respective year. The delta is colored in red if the absolute value is larger than 20 percent. "Same styles" only includes 110 unique styles which are produced in Units 1 and 2 and Unit 6, whereas "All styles" includes all styles regardless of how many units they are produced in.

high-performing Units 1 and 2 and the low-performing Unit 6, despite other external and internal factors being similar throughout. These two features of the data raise the questions of (1) how the factory managed to recover and (2) what differences in performance across units can tell us about the quality of management decisions.

3.4 Empirical strategy

To estimate the effects of external and internal production determinants on productivity- and quality-related outcomes, we define the following OLS specification, where the first row displays external determinants, the second row internal determinants related to style transitions, the third row internal determinants related to 'cumulative knowledge' at the line and unit level (which increases with

each additional day a style is produced on the line or in the unit within and across production spells), and the fourth row internal determinants related to 'working knowledge' at the line level (which increases for each style within a given production spell):

$$
\begin{aligned}
Y_{s,l,u,f,t} = {} & \beta_0 + \beta_1 smv_{s,f,t} + \beta_1 firstday_{s,l,f,t} \\
& + \beta_3 sd_smv_{l,f,t} + \beta_4 nb_styles_{l,f,t} + \beta_5 change_smv_{l,f,t} \\
& + \beta_6 run_days_line_{s,l,f,t} + \beta_7 run_days_unit_{s,u,f,t} \\
& + \sum_{i=1}^{6} \beta_{8i} run_days_capped_i_{s,l,f,t} + \sum_{i=2}^{6} \beta_{9i} capped_days_passed_i_{s,l,f,t} \\
& + \delta_f + \delta_b + \delta_t + \epsilon_{f,t}
\end{aligned}
\tag{3.1}
$$

Each variable included in the above specification is described in detail in the following. $Y_{s,l,u,f,t}$ is the outcome variable of interest for a specific style s on a particular line l under unit u in factory f at day t. Then, $smv_{s,f,t}$ is the style-specific standard minute value proxying the complexity of a style. $firstday_{s,l,f,t}$ is a dummy indicator equal to 1 for the first time a style is produced on a given line, 0 otherwise (called "First day style is produced on line X" in output tables). $smv_{s,f,t}$ and $firstday_{s,l,f,t}$ are 'external determinants', which cannot be changed by the factory on a daily basis and are shaped by buyers' demands. All other determinants included in the specification are 'internal determinants', which the factory has the discretion to change. First, we include three variables to track style transitions. Two of these measure the effects of producing multiple styles on the same line-day: $nb_styles_{l,f,t}$ is the number of styles produced on a given line, and $sd_smv_{l,f,t}$ is the standard deviation of the standard minute values of all styles produced on a given line per day (called "Number of styles produced each day by line" and "SD of SMV", respectively, in output tables). The third variable measures the effects of transitioning from one day to the next: $change_smv_{l,f,t}$ is the proportional change in the day average SMV on a given line from time t-1 to time t (called "Proportional change in SMV day average").

Next, we include variables to measure two types of learning-by-doing effects: First, the factory has the choice to produce the same style for an additional day per line within a production spell on a given line, or finish the production spell

on a different line. The former decision entails positive learning effects, which we describe as coming from more 'working knowledge', i.e. more coordinated processes, better functioning machines, higher availability of accessories, and workers getting better at what they are doing over time. Second, the factory has the choice to place styles on the same line or in the same unit at a later point in time again, building 'cumulative knowledge'.

To measure 'working knowledge' effects, we introduce two variables $run_days_capped_i_{s,l,f,t}$ (called "Running days (capped at 7)") and $capped_days_passed_i_{s,l,f,t}$ (called "Capped Days (6) passed"). $Run_days_capped_i_{s,l,f,t}$ is a running count of the number of days the same style is produced on a line by time t, where we reset the count to 1 whenever the time between two subsequent production entries for the same style on the line, i.e. the days passed, is a week or more. We censor the variable at 7 days, as the largest efficiency hits occur during the ramp-up time, i.e. during the first days a new style is produced per line. We introduce dummies and set the 7+ category as the reference group to measure the penalty associated with producing a style for one day shorter. Whenever we see two subsequent production entries for the same style on a given line with less than a week in between, it is very likely that the order has not been completed yet and the production has just been halted (e.g. due to missing inputs). In these situations, we do not reset running days, as we think it plausible to assume that there is no loss of style-dependent knowledge from such interruptions. To account for short breaks in producing the same style on the line, we introduce the variable $capped_days_passed_i_{s,l,f,t}$, which is a running count of the number of days between two (sorted) production entries for the same style on a given line and day. This variable is equal to zero the first time a style is produced on the line and is censored above at 6. This means that once a given style has been produced on the line, the variable equals the number of days passed between two subsequent production entries for the same style on the same line, up to a maximum of 6 days. Due to the inclusion of the dummy $firstday_{s,l,f,t}$, we set the base category equal to 1, and estimate the penalties associated with each additional day without producing the style up to the 6 days. If production for a given style is interrupted for more than six days, the penalty from prolonging the

break is captured by the reset of the running days variable, as described above.

To study the role of 'cumulative knowledge', we introduce two measures of cumulative experience to check separately for the presence of learning effects from an additional day of experience on style s at both the line and unit level. Specifically, $run_days_line_{s,u,f,t}$ is the running count of the days a specific style has been produced on a given line by time t, while $run_days_unit_{s,u,f,t}$ is the running count of the days a specific style has been produced in a particular unit by time t (called "Cumulative running days - Line" and "Cumulative running days - Unit", respectively).

In the regressions for our main factory of interest and the comparison factory we include month/year and buyer fixed effects. For robustness, we also ran two specifications where we include line fixed effects, and where we cluster standard errors at the style level. Further, we run the standard specification on a yearly basis, and for Units 1 and 2, and Unit 6 separately, allowing us to compare the magnitude of coefficients.

3.5 Results

3.5.1 Effects of external and internal determinants on outcomes

Table 3.4 shows regression results of our standard specification for the main factory. The regression includes month and year fixed effects and buyer fixed effects, but the results remain largely unchanged with line fixed effects (Table 3.B.5) or with both line fixed effects and standard errors clustered at the style level (Table 3.B.6).

Unsurprisingly, we find that producing more complex products is negatively associated with efficiency, with one additional minute of SMV linked to a 0.9 percentage point (roughly a 1.8 percent) decrease in efficiency (see coefficient on "Style SMV" in Table 3.4). Additionally, producing a style for the first time on the line leads to a 2.5 percentage point decrease (see coefficient on "First day style is produced on line X" in Table 3.4).

A noticeable set of results is the negative impact of style transitions within and across days. For instance, introducing one additional style is associated with a 1.93 percentage point decrease in efficiency (see coefficient on "Number of styles

produced each day by line" in Table 3.4). Furthermore, increasing the standard deviation of the SMV of the styles by one unit leads to a 0.7 percentage point decrease in efficiency (see coefficient on "SD of SMV" in Table 3.4). Our measure of variability across days for SMV, the proportional change in the mean SMV value from one day to the next (see coefficient on "Proportional change in SMV day average"), also shows large efficiency penalties, suggesting that a line's productivity suffers greatly when style turnover requires machines or operators to be moved, or entails operators having to work on processes that differ substantially from those produced the previous day. The coefficient is large but it must be standardized for a more meaningful interpretation, as the variation in the proportional change in the SMV average is relatively low. An increase of one standard deviation of this variable leads to a 1.85 percentage point decrease in efficiency - a more modest effect, but comparable to the disruption generated by producing an additional style on a given day.

In our exploration of learning effects, we distinguish between 'working knowledge' at the line level (which increases for each style within a given production spell) and 'cumulative knowledge' at the line or unit level (which increases with each additional day a style is produced on the line or in the unit within and across production spells). With regards to the former, we find that penalties associated with producing a style for one day shorter, i.e. the efficiency loss from foregone learning, are largest on the very first days of producing a new style relative to a scenario where the style stays on the line for a week or more (7+ as a reference group). In fact, the biggest hits to efficiency come from cutting production short on the first and second day, with a drop in efficiency of 7.45 percentage point and 3.10 percentage point respectively (see coefficients on "Running days (capped at 7)=1 and Running days (capped at 7)=2 in Table 3.4). The dissipation of knowledge happens mostly in the first three days, with the learning penalties settling at around half a percentage point after that, suggesting a faster increase in 'working knowledge' during the ramp-up time. We further find that interrupting production of a given style for a couple of days (less than a week) leads to additional and substantial penalties in efficiency, which vary by the duration of the break (see coefficients on "Capped days (6) passed in Table 3.4). The penalty for interrupting production

for two days relative to the shortest possible break of one day is roughly a 6 percentage point decrease in efficiency. The penalties gradually increase and peak at a break spell of four days, where the efficiency drop when production is taken up again is of almost 9 percentage points. While the penalty decreases to 4.9 percentage points for breaks of six or more days, we argue that this is only the residual effect that is not picked by the complete loss of working knowledge as the running days are reset to 1.

Turning to 'cumulative knowledge' effects, we find that an additional day producing a given style on the line overall is associated with a 0.135 percent increase in efficiency, as indicated by the significantly positive coefficient on "Cumulative running days - line" in Table 3.4). This suggests that the line's "stock" of style-dependent knowledge improves the line's performance when producing that style. Unexpectedly, we find that the same measure at the unit level is negatively associated with performance. However, the coefficient on "Cumulative running days - Unit" in Table 3.4 is not economically meaningful and only significant at the 10 percent level.

Finally, we see that the factory performs better for some buyers than others, as indicated by significant coefficients on some of the buyer indicator variables. When revealing the identity of these buyers (kept anonymous in this paper), we can see that they are particularly important customers with larger order sizes. This suggests that management pays closer attention to orders from these customers to ensure a particularly efficient and flawless production.

The patterns in the two quality variables are broadly consistent with the ones for efficiency. This is particularly true for the alteration rate measure. Of course, the signs are all reversed, because an increase in efficiency and a decrease in alteration (or rejection) rates are both positive outcomes. We see that the signs are indeed reversed for the alteration rates throughout, though alteration rates are no higher on the first day a style is produced on the line, or when multiple styles are produced on a day. Moreover, gaps in production (days passed) appear to have no effect on quality. But alteration rates are significantly increasing in complexity, the variance in the SMVs of the styles being produced on the day, and changeovers to styles with different SMV. Moreover, alteration rates are significantly higher over

the first four days a style is produced. As with efficiency, the cumulative learning measure at the unit level has an unexpected sign and is highly significant for both alterations and rejects. This is a puzzling result in regressions that otherwise yield expected patterns.

TABLE 3.4. STANDARD SPECIFICATION

	Efficiency	Alteration Rate	Rejection Rate
	(1)	(2)	(3)
Style SMV	-0.852***	0.367***	0.020***
	(0.000)	(0.000)	(0.000)
First day style is produced on line X	-2.506***	0.021	0.007
	(0.000)	(0.729)	(0.604)
SD of SMV (line level)	-0.722***	0.040***	0.013***
	(0.000)	(0.008)	(0.000)
Number of styles produced each day by line	-1.928***	-0.001	0.036***
	(0.000)	(0.926)	(0.000)
Proportional change in SMV day average	-15.259***	0.983***	0.033
	(0.000)	(0.000)	(0.161)
Cumulative running days - Line	0.135***	-0.030***	-0.001
	(0.000)	(0.000)	(0.257)
Cumulative running days - Unit	-0.009*	0.008***	0.001***
	(0.072)	(0.000)	(0.001)
Running days (capped at 7)=1	-7.446***	0.475***	0.015
	(0.000)	(0.000)	(0.343)
Running days (capped at 7)=2	-3.101***	0.426***	-0.016*
	(0.000)	(0.000)	(0.085)
Running days (capped at 7)=3	-1.508***	0.210***	-0.011
	(0.000)	(0.000)	(0.254)
Running days (capped at 7)=4	-0.578***	0.085**	-0.020**
	(0.001)	(0.020)	(0.028)
Running days (capped at 7)=5	-0.547***	0.050	-0.011
	(0.003)	(0.187)	(0.202)
Running days (capped at 7)=6	-0.630***	0.071*	-0.001
	(0.001)	(0.072)	(0.904)
Running days (capped at 7)=7	0.000	0.000	0.000
	(.)	(.)	(.)
Capped Days (6) passed=1	0.000	0.000	0.000
	(.)	(.)	(.)
Capped Days (6) passed=2	-6.043***	0.064	-0.002
	(0.000)	(0.136)	(0.860)
Capped Days (6) passed=3	-7.965***	-0.014	0.005
	(0.000)	(0.833)	(0.709)
Capped Days (6) passed=4	-8.931***	-0.022	-0.003
	(0.000)	(0.789)	(0.893)
Capped Days (6) passed=5	-8.810***	0.139	0.001
	(0.000)	(0.165)	(0.965)
Capped Days (6) passed=6	-4.907***	-0.277***	-0.002
	(0.000)	(0.000)	(0.894)
Constant	38.756***	5.578***	0.411***
	(0.000)	(0.000)	(0.000)
RD (line)=RD (unit)	0.00	0.00	0.07
Line FE	No	No	No
Month/year FE	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes

Notes: Running days are reset at days passed=7, capped at 7+, and have 7+ as the base category. Days passed is capped at 6+, with 1 as the base category). We use robust standard errors and allow for month/year, and buyer fixed effects. * p<0.1, ** p<0.05, *** p<0.01

3.5.2 Comparison of effects to the industry

Regression results of our standard specification for the comparison factory are shown in appendix Table 3.B.10. Comparing the size of the coefficients to our main factory, we can see that effects from internal determinants related to style transitions are much larger in the comparison factory. This is not surprising given that our main factory experiences many more style transitions overall than the comparison factory (see summary statistics in Table 3.1). It suggests that the main factory is better at coping with style transitions, a result that we believe could be driven by the nature of the changes to the factory's production model, which entailed faster and more frequent style transitions. On the contrary, inefficiencies caused by the ramp-up of new styles, particularly on the first day, are much larger in the main factory than in the industry as shown in Figure 3.5. We interpret this as a sign that our main factory seems to accumulate working knowledge at a faster pace than the industry factory, particularly in the first few days of a production spell, something that would be consistent with more experience with constant style transitions. An alternative explanation could be that the main factory is not as good at starting production of new styles because middle management is not able to dedicate sufficient attention to the set-up of a new style due to the high number of styles being produced per day (relative to the industry), leading to particularly harder hits when the style is (re)introduced on the line. Yearly regressions in Table 3.B.7 show that only 2016's coefficient size is noticeably different from other years, with the pattern being relatively stable otherwise.

3.5.3 Comparison of effects across years and units

When comparing magnitudes of the coefficients across years, we pay particular attention to anomalies in 2017, the year in which efficiency reached its lowest level. The most notable difference to preceding years is the change in the sign of the coefficients on the number of styles produced each day by line, which changed from +3.86 in 2016 to -1.5 in 2017. This coincides with a sharp increase in the actual number of styles produced each day by line. This is in line with anecdotal evidence suggesting a conscious management decision by the CEO to change the factory's

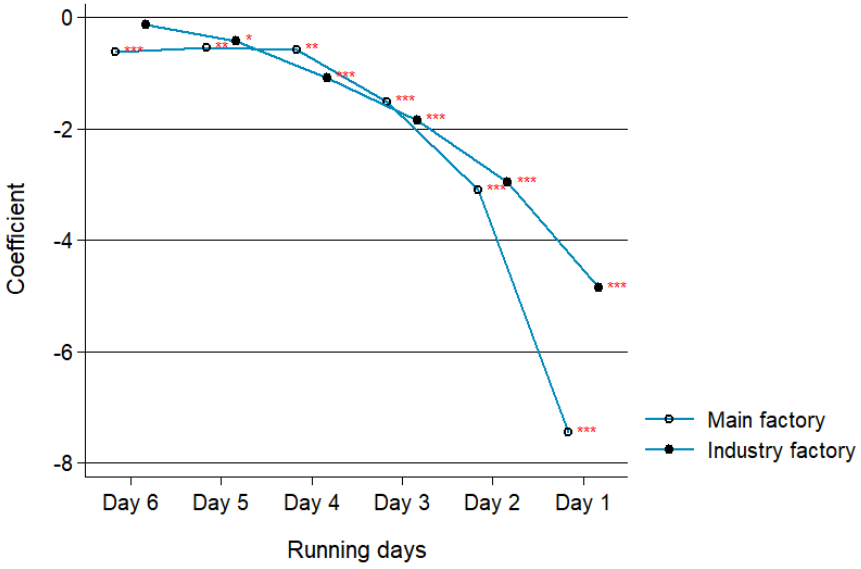


Figure 3.5: Comparison of effects of working knowledge

Notes: Figure 3.5 compares the coefficients on running days of the regressions in Table 3.4 and Table 3.B.10, with efficiency as the dependent variable. * p<0.1, ** p<0.05, *** p<0.01

production model to adapt to new trends in 2017, as we qualitatively described in Section 3.3.

We do not see any remarkable differences in the magnitudes of the effects between Units 1 and 2, and Unit 6, as shown in Appendix Table 3.B.11. If anything, the well-performing Units 1 and 2 have more negative coefficients than Unit 6, which suggests that we need to go beyond a comparison of the magnitude of these effects to explain performance differences across units.

3.6 Learning and autonomy within organisational entities

Industrial Engineers in the planning department are responsible for designing a line layout for each style that the factory will produce. They decide how to break the assembly of the garment into individual processes, what machines to use for each step, and how many pieces will be produced each hour by an operator working on each process. Together with the order delivery schedule, these data are key inputs used in generating the monthly production plans. In this section, we identify deviations from the production plans in the actual production data to shed

light on learning processes across the planning and production departments, and autonomy within production units, i.e. the "real authority" (Aghion and Tirole, 1997) exercised by the production managers in each unit. We focus on deviations from the plan solely arising from the production department.[14]

We first qualitatively analyse how well production managers stick to plans. Specifically, we check whether styles are produced on the exact day they were planned, either on the planned line or on other lines within the original line's organisational group.[15] We then use the regression coefficients described above to calculate the efficiency cost from variables related to style turnover embedded into the production plan, and compare those to the efficiency costs from those same factors actually incurred during production.

3.6.1 Qualitative description of deviations from plans

Managers are faced with the choice of sticking to the production plans provided by the production department, or changing the allocation of styles and production quantities on a day-to-day basis across the various lines. In the first subsection, we analyse the percentage of styles that are produced on the line and day called for in the plan. We categorise correctly planned styles as styles that are produced on the exact day and line they were planned. We also look at the number of styles that are actually produced compared to the planned number. In the second subsection, we look at the percentage of styles produced on the planned day on *any* production line within the same group rather than the exact planned line.

To do so, we match styles in the planning data with those in the production data at the line and day level and check for each day whether the styles planned for that date appear in the production data, and see how the accuracy holds up for different planning horizons. The different planning horizons are measured by the

[14]Deviations may arise from (i) the production department's inability to adjust to new industry demands, (ii) because of last minute changes in the orders that could not be reflected in the planning data, or (iii) departments other than production (e.g., the fabric production unit) failing to keep to plan. We rule out (ii) because buyers place orders well in advance and there is no evidence (in our datasets or anecdotally) pointing in that direction. We aim to quantitatively rule out (iii) with additional data in a future version of the paper.

[15]A 'group' is the combination of three to four units into one organisational entity. Units 1 and 2 belong to Group 1, which consist of four units in total. Unit 6 belongs to Group 2, which consists of three units in total.

days between the production date and the week the plan was made. We take the first day of the respective week, which explains why the planning horizon in days starts at 7 days. Most of the plans are made for 30 consecutive days, though some cover 60 days. Matching the planning and production data was not straightforward because the two departments use different names for the same style. Thus, the task required extensive data cleaning. Subsection 3..1 details the steps taken and the resulting match rates. Overall, we were able to match roughly 61 percent of styles shown in the production data to the same style in the planning data.[16] This allowed us to create a bridge between production and planning for 52 percent of the observations in the production data. There are many instances where a single style in the planning data refers to multiple styles in the production data, and many where the reverse is true. As depicted in Table 3..3, match rates are similar across years and units. Overall, the matched and unmatched styles do not differ substantially from one another as shown in Table 3..2. For this reason, we do not believe there are systematic reasons behind the failure to match. The resulting set of matched styles is very large and, we believe, reasonably representative.

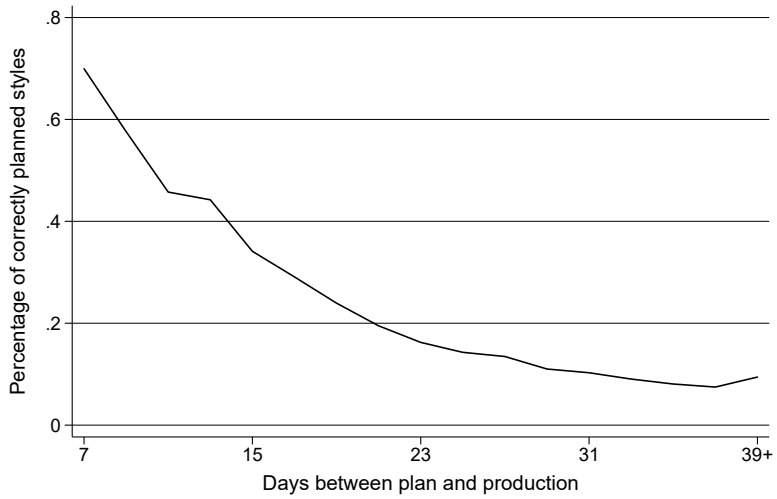


Figure 3.6: Percentage of correctly planned styles by planning horizon

Notes: Figure 3.6 shows the percentage of correctly planned styles by planning horizon. The planning horizon is the number of days between the production date and the week the plan was made. We take the first day of the respective week, and thus the x-axis starts at 7 days.

[16]This is a weighted average across all years using the number of unique styles in production each year as weights.

Deviations from producing on exact planned lines

We first look at the percentage of planned styles which were accurately produced on the planned line on the planned day. We show this percentage for different planning horizons in Figure 3.6. The share of correctly planned styles decreases the further away the date of the plan is from the actual production date. Plans made in the week before the actual production date are on average 70 percent accurate, whereas plans made three weeks before are only around 20 percent accurate.

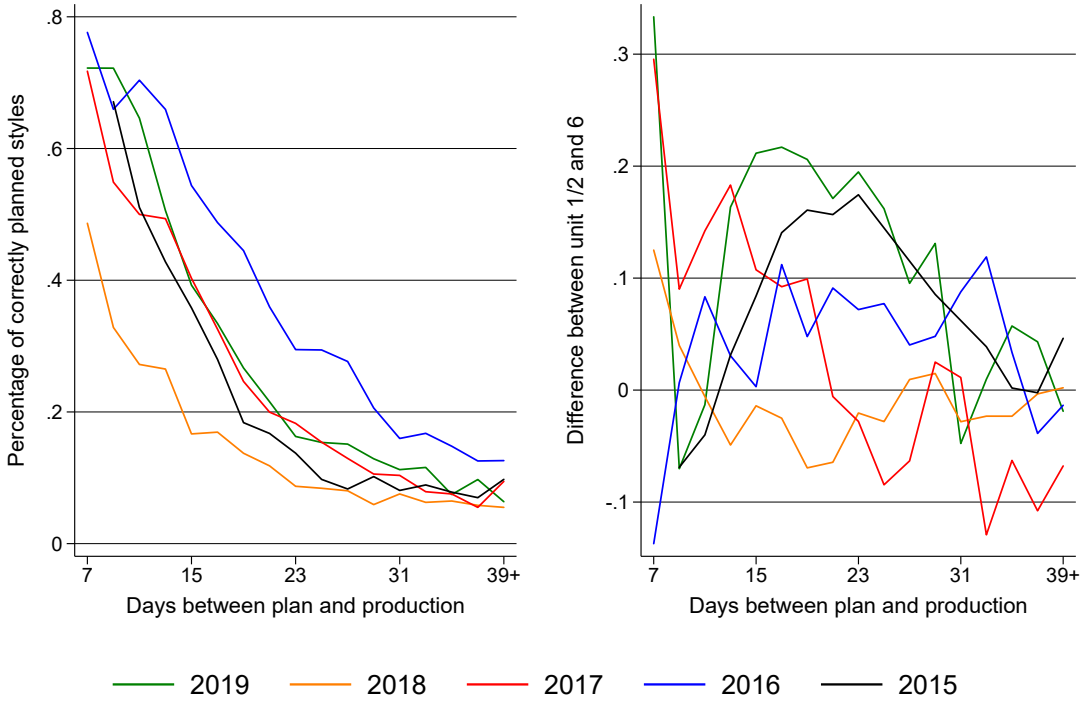


Figure 3.7: Percentage of correctly planned styles by planning horizon per year

Notes: The left graph of Figure 3.7 shows the percentage of "correctly" planned styles by planning horizon across years. The right graph of shows the proportional difference in the share of styles produced on the exact day and line they were planned between Units 1 and 2, and Unit 6 six per year. The planning horizon is the number of days between the production date and the week the plan was made. We take the first day of the respective week, and thus the x-axis starts at 7 days.

A comparison of the percentage of correctly planned styles across years in the left graph of Figure 3.7 reveals remarkable differences in the production managers' decisions or lines' ability to stick to plans. In 2015, the accuracy of plans is reasonably high, with an increase in 2016 just before the factory changed its production

model. At first, when the change takes place in 2017, the accuracy does not drop below the 2015 level. Instead we see a sharp decrease in accuracy in 2018, followed by a recovery from 2019 on-wards.

We compare the higher-performing Units 1 and 2 with the lower-performing Unit 6 and show the proportional difference in the share of styles produced on the exact day and line they were planned between Units 1 and 2, and Unit 6 per year in the right graph of Figure 3.7 (see Appendix Figure 3.A.5 for separate shares in Units 1 and 2, and Unit 6). While the accuracy is generally larger in higher-performing units than in the lower-performing unit (see difference larger than zero in Figure 3.7), it is actually slightly *lower* in 2018 for styles planned more than 10 days in advance. The same pattern holds for the number of styles to be produced on a given day per production line.[17] As shown in the left graph of Figure 3.8, the

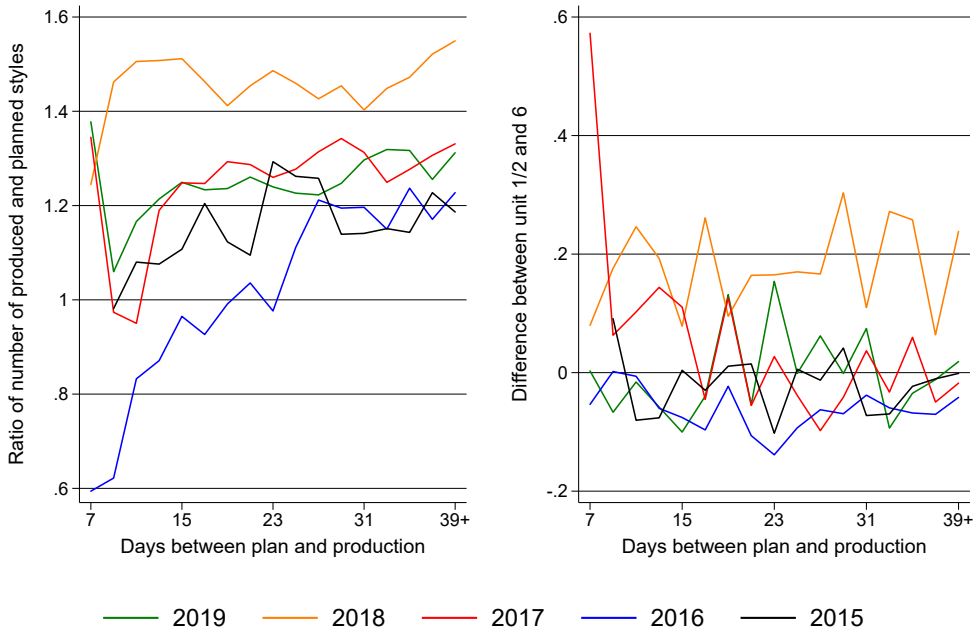


Figure 3.8: Ratio of number of produced and planned styles

Notes: The left graph of Figure 3.8 shows the ratio of the number of styles per line-day in the production data and in the planning data per year. The right graph shows the difference in the ratio between Unit 1 and 2, and Unit 6.

ratio of the produced number of styles to the planned number of styles is much

[17]For this analysis, we use the full planning and production data, i.e. matched and unmatched data, as the number of styles in both datasets can be calculated independently.

larger in 2018 than in other years. We interpret this as evidence that the factory largely deviated from the plan to allow more styles to be produced on a given day per line in 2018. The deviation is stronger in the high-performing Units 1 and 2, than in low-performing Unit 6 (see right graph of Figure 3.8).

Overall, we find that the factory as a whole generally struggles to plan accurately for longer planning horizons, and deviates substantially more from the plan in 2018 than in other years. We also find that higher-performing units deviate *more* from plans in 2018 compared to lower-performing units.

Deviations from producing on any production line

We now look at the percentage of correctly planned styles at the right day on *any* line within the same group rather than the exact planned line.

As expected, the accuracy now increases. The accuracy is only slightly higher for very short planning horizons, but it shows a much less sharp decrease at longer planning horizons, staying almost entirely above 40 percent (see upper graph of Appendix Figure 3.A.7). This pattern holds across all years except 2015 as shown in the left graph of Figure 3.9. This suggests that exercising autonomy and allowing additional flexibility to produce within the same group better enables the factory to keep up with a higher number of orders each day, even for longer planning horizons. The accuracy remains roughly unchanged if we allow for re-shuffling across the whole factory, suggesting the flexibility in production stays within organisational groups (see lower graph of Appendix Figure 3.A.7).

When looking at the difference in accuracy between the higher-performing units and the lower-performing unit in the right graph of Figure 3.9, we can see that the accuracy is now substantially larger in the higher-performing units compared to the lower-performing unit, even in 2018 - when it was lower if we restricted the definition of accuracy to the exact same lines (see Appendix Figure 3.A.6 for separate shares in Units 1 and 2, and Unit 6). We interpret this as evidence for the higher-performing units exercising more "real authority" than the lower-performing units, as shown by the way they flexibly re-allocate styles on a given day to different production lines within the same group to fulfil their targets.

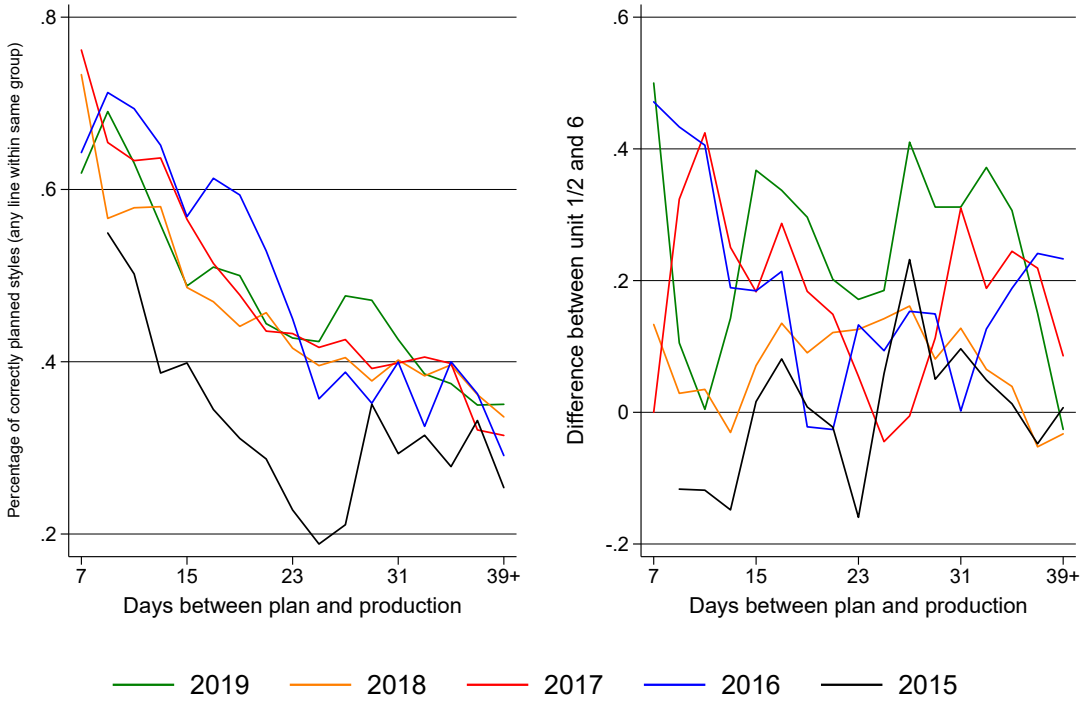


Figure 3.9: Percentage of correctly planned styles on any production line (within the same group) by planning horizon per year

Notes: The left graph of Figure 3.9 shows the percentage of "correctly" planned styles on *any* production line within the same organisational group by planning horizon across years. The right graph shows the difference in the share of correctly planned styles on *any* production line within the same organisational group at a given day by planning horizon between units one and two, and unit six per year. The planning horizon is the number of days between the production date and the week the plan was made. We take the first day of the respective week, and thus the x-axis starts at 7 days.

The re-shuffling of planned styles at the group level raises the question of how much production units within one group coordinate with one another. Our hypothesis is that re-shuffling at the group level requires a significant amount of coordination and communication across units within that 'autonomous' group. This would show up in the data by styles being produced in other units *in addition* to the originally planned unit rather than being produced *only* in other units. Figure 3.10 shows the share of correctly planned styles on any production line for each possible combination of production locations within the same organisational group. Correctly planned styles within the group can either be produced on a line within

the same unit, on a line in another unit, but in the same group, or simultaneously on a line in the same unit and on a line in another unit within the same group. As before, we categorise correctly planned styles as styles that are produced on the exact day it was planned. Figure 3.10 clearly indicates that a majority of the styles are produced in other units in addition to Units 1 and 2, whereas most of the correctly planned styles within the same group in Unit 6 are only produced in that unit. This could imply more coordinated decision-making and communication within those organisational groups to whom the higher-performing units belong, breaking down silo mentalities.

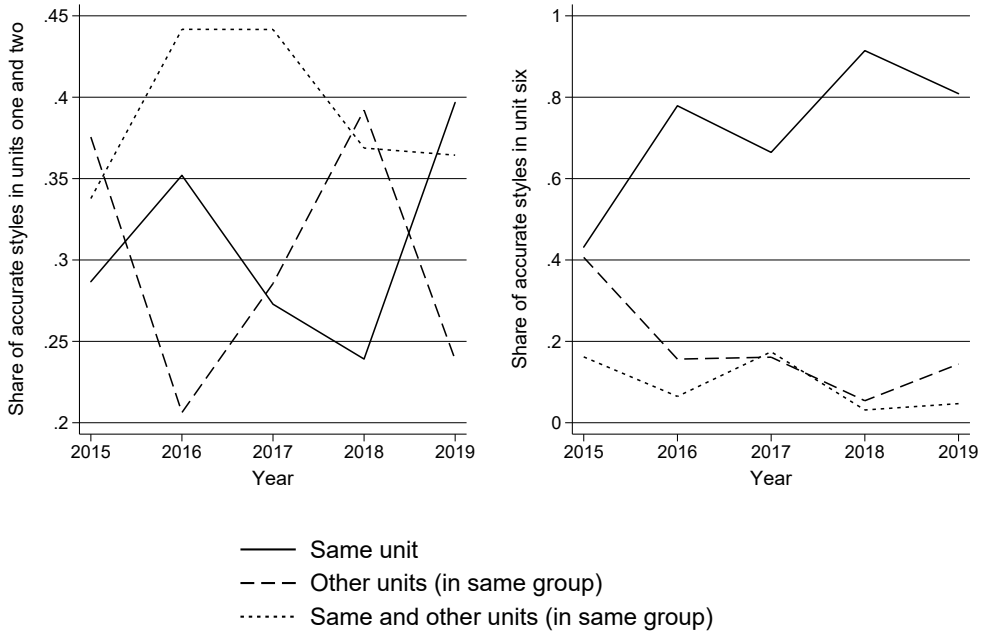


Figure 3.10: Production locations of correctly planned styles

Notes: Figure 3.10 shows the share of correctly planned styles on any production line for each possible combination of production locations within the same organisational group. Correctly planned styles within the group can either be produced on a line within the same unit ("Same unit"), on a line in another unit, but in the same group ("Other units (in same group)") or simultaneously on a line in the same unit and on a line in another unit within the same group ("Same and other units(in same group)"). As before, we categorise correctly planned styles as styles that are produced on the exact day they were planned. The left graph shows the shares for Units 1 and 2, and the right graph for Unit 6.

Overall, we conclude that the accuracy of plans is decreasing in the length of the planning horizon. Unexpectedly, the drop in accuracy does not come in 2017, the year of the changed production model, but one year after in 2018. This is the

case for the accuracy of producing both the correct styles and the correct number of styles per line on a given day. Further, we find that the accuracy of plans is clearly higher in the higher-performing Units 1 and 2 than in lower-performing Unit 6 when we relax the definition of accuracy to include styles produced within the same organisational group as well. In fact, in higher-performing units, most styles are actually produced in other units *in addition* to the same unit to ensure timely and efficient production of required styles. All together, we interpret this as evidence for higher autonomy within higher-performing production units. Production managers in those units seem to exercise their "real authority" (Aghion and Tirole, 1997) by deviating from plans and simultaneously coordinating with other production units within their same organisational group.

3.6.2 Costs of efficiency

To validate and elaborate upon the findings of the previous section that not sticking to plans might be a good management decision, we calculate efficiency costs embedded into the production plan, and compare those to the actual efficiency costs incurred during production. In doing so, we focus on three types of efficiency costs that stem solely from internal determinants related to style transitions, and thus are controllable by managers on a day-to-day basis. Namely, we consider costs associated with (i) producing one additional style per line on a given day, (ii) producing less similar styles on a given day, and (iii) producing less similar styles on a line from one day to the next. To estimate the difference in efficiency costs, we first store the magnitude of the coefficients on these three variables from the standard regression as shown in Table 3.4. We then draw upon detailed production and planning line-level data to apply these coefficients and obtain an estimate of the inefficiency in actual production as opposed to inefficiency embedded into the production plan.

Figure 3.11 shows the style transition-related inefficiencies embedded into the production plan and in the actual production, jointly with the evolution of efficiency across time. Up until the end of 2016, the factory seems to make conservative plans with higher inefficiencies embedded into the plan than in reality. In early 2017, when the blue and brown lines cross, inefficiency in production tends

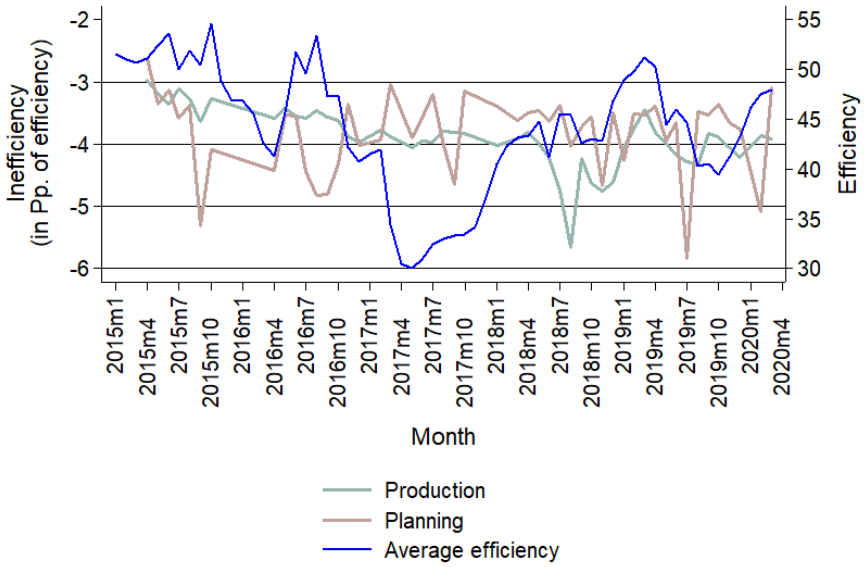


Figure 3.11: Efficiency costs in production and plan

Notes: Figure 3.11 shows efficiency costs from factors related to style transitions embedded into the plan in brown, those from actual production in green, and the trend in average efficiency in blue. Calculations include all units, and use coefficients related to style transitions from the efficiency regression in Table 3.4.

to stay below that embedded into the plan, but generally very close. The production managers try to stick very closely to plans when the factory's production model imposed by industry demand changes in 2017, coinciding with the plunge in efficiency. Around October 2017, the gap between inefficiencies widens up.

This development is also illustrated in Figure 3.12, which shows that the delta in efficiency costs between production and plan are largest in 2018. It seems as if production managers make a deliberate decision to relax their adherence to the plans to allow for greater flexibility in production, which implies higher efficiency costs in production than in the plan from style turnovers, but is associated with overall efficiency increases that take the factory to levels preceding the shock. We interpret this development as evidence that the planning department did not adapt sufficiently quickly to changes in the factory's production model, which would have instead required plans to account for many more styles to be produced on a given line for shorter periods of time. The production department seems to have given up on the plans to improve their ability to meet their targets. This could also suggest that the exchange of information and feedback between the planning

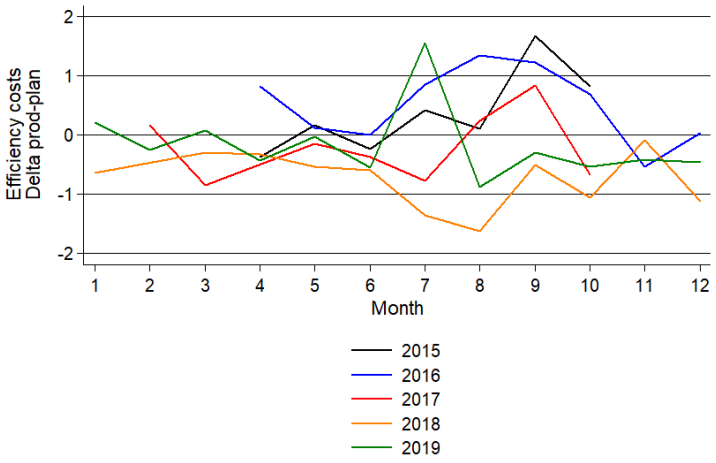


Figure 3.12: Delta in efficiency costs in production and plan across years

Notes: Figure 3.12 shows the difference in efficiency costs from factors related to style transitions between actual production and planned production ($production - plan$) for each year. Calculations include all units, and use coefficients related to style transitions from the efficiency regression in Table 3.4.

and production department was not sufficient. Only around early 2019, the gap between efficiency costs in production and plan closes, followed by some minor fluctuations towards the end of the year. At the same time, we can see efficiency continue its recovery to the 2015 level. The graph suggests that the planning department has learned from the production department and caught up on the idea that production has to be more flexible in terms of the number and location of produced styles, and updated its planning strategy.

A comparison of the difference in efficiency costs between production and plan in the lower-performing Unit 6 and the higher-performing Units 1 and 2 in Figure 3.13 shows that the 'crossing point', when the production department appears to deviate from the plans to allow for additional flexibility, occurs much later in Unit 6 compared to in Units 1 and 2. This could be pointing to a slower realization by the production department in Unit 6 that plans had become inadequate following the changes to the production model, and could also suggest a lack of communication of production departments across units. The production department in Unit 6 seems to continue sticking to production plans much longer and allow for less flexibility in production than Units 1 and 2, coming at the cost of the larger efficiency decrease outlined in Section 3.3.

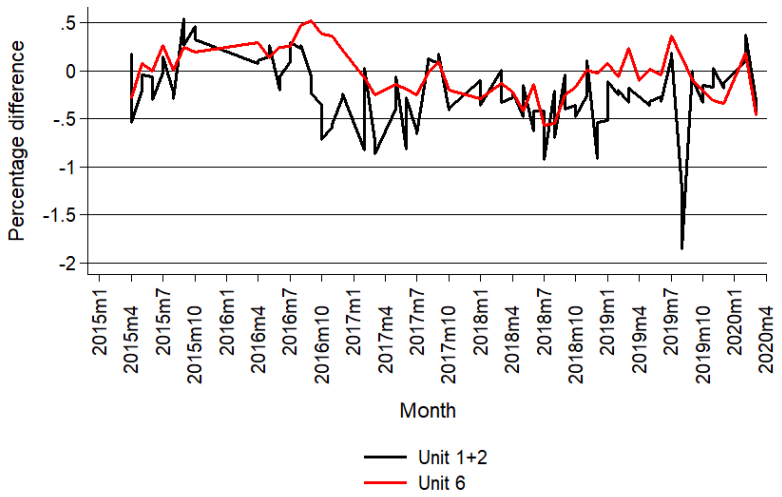


Figure 3.13: Differences in efficiency costs per unit

Notes: Figure 3.13 shows the proportional difference in efficiency costs from factors related to style transitions between planned production and actual production ($\frac{plan-production}{plan}$) for Units 1 and 2 (black line) and for Unit 6 (red line). Calculations use coefficients related to style transitions from the efficiency regression in Table 3.4.

Next, we decompose efficiency costs into its three components: (i) producing one additional style per line on a given day ("Loss from multiple styles"), (ii) increasing the variation in style complexities per line on a given day ("Loss from variation in daily style complexity"), and (iii) increasing the variation in style complexities per line from one day to the next ("Loss from style transitions"). The left panel of Figure 3.14 shows that losses from multiple styles are biggest in the actual production, despite the regression coefficient being only -1.9 (as depicted in Table 3.4). Losses from variation in the daily style complexity are the lowest, in line with a small regression coefficient of -0.7. Losses from style transitions are also reasonable low because, while the regression coefficient was largely negative (-15.3 percentage points), the variation in the variable is quite low. A similar pattern holds for the decomposition of efficiency costs embedded into the production plan as shown by the right panel of Figure 3.14, though fluctuations are more volatile particularly in early 2017. This decomposition shows that increasing the number of styles per line have the most detrimental effects on efficiency. The factory could be better off by taking this (as well as the other two factors) into account to achieve a more optimal re-allocation of orders across production lines once deviating from

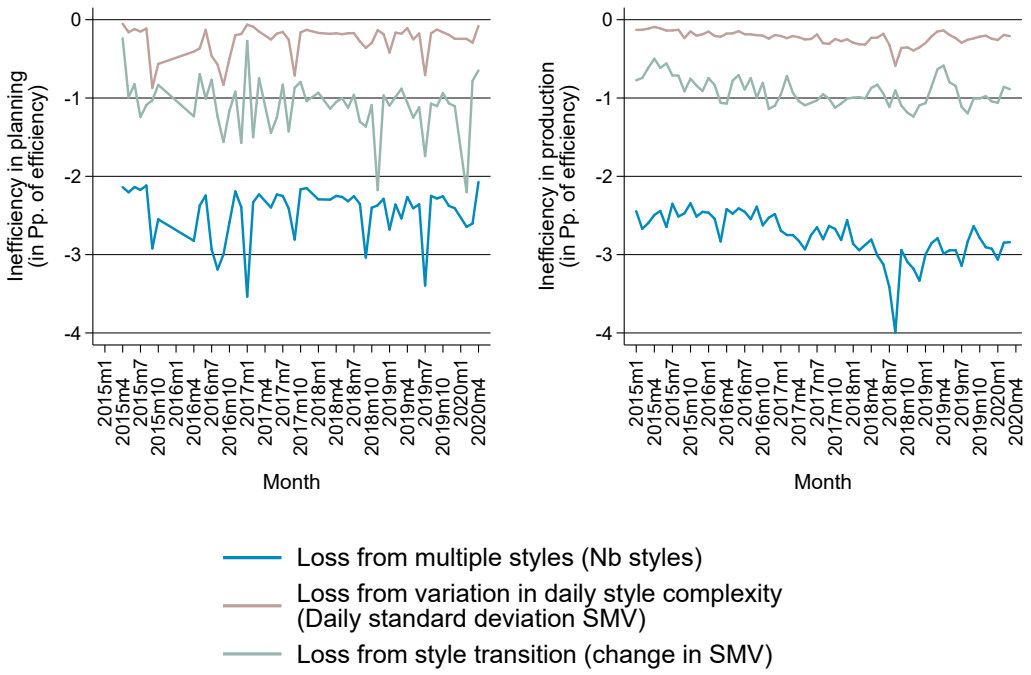


Figure 3.14: Decomposition of efficiency costs

Notes: Figure 3.14 shows the decomposition of efficiency costs into its three components: (i) Loss from multiple styles, (ii) Loss from variations in daily style complexity, and (iii) Loss from style transitions. Decompositions include all units and are based on regression coefficients of Table 3.4. The left graph shows the decomposition using the planning data, and the right graph using the production data.

the plan becomes unavoidable.

We conclude that the production department first attempts to adhere to the new planning strategy after the change in the production model, but then deviates from plans to keep up with the high-pace of the industry despite introducing higher efficiency costs from factors related to style transitions. Only when the planning department learns from the production department and updates their planning strategy, the gap between production and planning closes. This pattern is more salient in higher-performing Units 1 and 2, which allow for a more flexible approach to production earlier on compared to the lower-performing Unit 6 and hence, better manage to meet their targets as measured by their higher efficiency.

3.7 Conclusion

Firms operating in low-trust and highly volatile business environments are faced with the tension between a desire to centralise decision-making in response to low levels of trust, and the need for flexibility to deal with frequent shocks that arise from the more volatile business environments. In this paper, we analyse detailed line-level data from 118 production lines in a large Bangladeshi garment factory to examine this tension. We first describe trends in the data, in particular (i) a large efficiency decrease in 2017 caused by switching to a 'fast fashion' production model imposed by industry demand, (ii) a subsequent impressive, but uneven, recovery despite unchanged conditions. We attempt to understand how the factory managed to recover despite additional changes to its production model, and why performance differences across units persisted. Part of the explanation for the first question lies in the presence of learning effects. First, the factory appears to be better at coping with style transitions than comparable factories. We believe that this is due to the nature of the changes to the factory's production model, which entailed faster and more frequent style transitions. For example, producing one additional style on a production line on a given day is associated with a 1.93 percent decrease in efficiency as compared to 3.4 percent in a comparison factory.

Second, we find positive effects from building 'working knowledge' at the line level and 'cumulative knowledge' at the line, but not at the unit level. The factory seems to accumulate working knowledge at a faster pace than the industry benchmark, as seen by the fact that our main factory suffers larger hits to efficiency from cutting production of a style short relative to prolonging the production spell to a week or more. For instance, keeping production of a style to just one day relative to producing it for a week or more is linked to a 7.45 percent decrease in our main factory, but only a 4.86 percent decrease in the comparison factory. The maintenance of the line's working knowledge is important in the context of fast and frequent transitions, as seen by the fact that each additional "break day" in the production of a style leads to large efficiency costs - effects that are once again larger than in the industry factory.

Third, we find that performance differences across units can be traced back to varying degrees of autonomy exercised by production managers in deviating

from pre-determined plans. After comparing the efficiency costs in the production data with those embedded into the plans, we find that the change in the factory's production model undermined the planning departments' ability to generate adequate production plans. The production department first attempts to adhere to the new planning strategy after the change in the production model, but then deviates from plans to keep up with the high pace of the industry despite higher efficiency costs from factors related to style transitions. Only when the planning department learns from the production department and updates their planning strategy, the gap between production and planning closes. Furthermore, we observe that this pattern is more pronounced in the higher-performing Units 1 and 2, and that this might explain the persistence of performance differences within the firm. The higher performing Units 1 and 2 seem to exercise more 'real' authority (Aghion and Tirole, 1997) by deviating from plans *more* than lower-performing units. In turn, this more flexible approach to production earlier appears to help the higher-performing units to ensure a more timely and efficient production of required styles.

In a future version of this paper, we attempt to provide further evidence to ratify the robustness of our results. First, further work should probe more deeply into other potential drivers of performance differences across units. For example, while we were able to rule out heterogeneous product mixes, we were not able to rule out the possibility that higher-performing units (i) hand over more troublesome styles to lower-performing units, (ii) get fabric first and hence, can be more productive earlier on, (iii) have a higher-skilled workforce and lower turnover.

Second, while we provided descriptive evidence of learning patterns and degrees of autonomy across time, further research is needed to sharpen this evidence using statistical tests. For example, one could test whether differences across years and units as outlined in Section 3.6 are statistically significant.

Third, more research can also be done on the pressure imposed by the proximity of delivery dates. On the one hand, it would be interesting to see whether high- and low-performing units vary in their capacities to meet delivery dates. On the other hand, our analysis would benefit from including a variable in the regressions with the time between the delivery date and the current production date (proxying

for production pressure) to analyse how lines' performance is affected when there is a smaller margin for error.

Fourth, additional research is needed to devise algorithms that allow to extend the comparisons of the production and the planning data to longer time horizons in production. For example, it would be interesting to analyse the days of delay until a style is actually produced on the planned line or on other lines in the same unit. Similarly, we think it would be fruitful to see how the share of "correctly planned styles" varies when we relax our definition of accuracy to allow for plans to be produced e.g. 5 or 10 days after the originally planned date.

An overall limitation of this study is owed to the factory's inconsistent data tracking across departments. While we were able to find a corresponding planning style for 52 percent of the production data, this percentage is still relatively low and, despite the lack of meaningful differences between matched and non matched observations, we would have liked to carry out a more comprehensive analysis. Ideally, future research will guide firms in the data tracking process in a prospective way to allow for analyses of the adequacy of their plans and the extent of the communication across departments in the firm. Finally, it would also be interesting to enrich the data with an assessment of managers' skills and see whether coordination processes between the planning and production department are correlated with managers' capabilities.

Acknowledgements We would like to thank the data team at the Innovations for Poverty (IPA) office in Bangladesh for their hard work on this project. Ethical approval was obtained on the 7th of August 2018 from the Oxford Department of International Development's Departmental Research Ethics Committee (DREC) with reference C1A_18_079.

Appendices of Chapter 3

3..1 Data handling

The factories in our sample often refer to the same style with slightly different style identifiers, an approach intended to account for minor changes in the features of the item or color variations. This introduced substantial difficulties in our efforts to accurately track lines' production of a given style. As the knowledge required to produce a given item is not dependent on the color of the cloth being used, the factories' approach to classifying their products led to a fictitiously high number of distinct styles in the data for the purpose of this analysis. To account for this, we implemented a strategy that groups identifiers that refer to (virtually) the same style. This was done through a two-stage process. First, we sorted the data on the basis of the style identifiers and created a first grouping level in which styles were assigned to the same group if (i) the Levenshtein ratio (Levenshtein, 1966) score based on the identifier strings between adjacent style entries is above 0.89,[18] (ii) the first two characters of the strings match,[19] and (iii) the order for the two styles was placed by the same buyer, after grouping together sub-brands from the same parent company. Second, we design and implement a kmeans clustering algorithm in STATA to create subgroups within each grouping level 1 based on the complexity of the relevant styles. In doing so, we require that the maximum SMV within the subgroup to be at most 10 percent larger than the minimum SMV, for all the subgroups created.[20]

An important piece to the productivity puzzle in our main factory is the extent to which actual production follows the production plans of the factory. A main obstacle in setting up this analysis was the fact that the factory used different style identifiers in the production and planning data. To bypass this issue, we created yearly style libraries across the two datasets to create a correspondence between production and planning styles. Our library yielded a 60.7 percent average match rate across all years (weighted by the number of unique styles in production each year, i.e. column "Syle N" in Table 3..1).

The matching process is carried out on a yearly basis in order to maximize the match rates each year. Specifically, we attempt to match each year of production data to its corresponding planning data. In doing so, we take a conservative approach to the selection of the planning data. On the one hand, the factory can

[18]The Levenshtein ratio is a measure of the differences between two strings. This ratio relies on the Levenshtein distance between the strings, which is defined as the sum of the penalties linked to the each step required to turn string b into string a (cost of substitution =2 , cost of insertion and deletion=1). We implement this using the fuzzywuzzy python module. Specifically, this module calculates the ratio score between strings a and b by (1) calculating the Levenshtein distance between the strings (i.e. the "penalty"), (2) calculating the sum of the lengths of the two strings minus the "penalty", and (3) dividing this quantity by the sum of the lengths of the strings.

[19]Two style identifiers referring to (virtually) the same style would never differ in its initial characters. We found that requiring this condition removed mistakes in the grouping algorithm that were not accounted for by the Levenshtein ratio score.

[20]Kmeans clustering minimizes within-cluster squared Euclidean distances. We implement this in such a way that, for each grouping level 1, the algorithm is re-run and the number of required clusters increased for as long as the all of the resulting subgroups do not pass the threshold test, where the threshold test requires the max/min ratio of SMVs within the subgroup to be below 1.1.

TABLE 3..1. MATCH RATES FROM STYLE LIBRARY

Year	Style N	Step 1	Step 2	Step 3	Step 4	Step 5	Match	Match %
2015	1281	172	171	142	0	0	**485**	37,86
2016	1756	281	330	422	2	0	**1035**	58,94
2017	2311	291	448	571	11	0	**1321**	57,16
2018	2624	216	960	602	119	0	**1897**	72,29
2019	2732	276	699	756	82	123	**1936**	70,86
2020	979	6	104	302	3	4	**419**	42,80

Notes: Table 3..1 displays the average match rate for the unique list of styles identifiers in each year in the production data yielded by the matching algorithm used to create a bridge between the production and planning style identifiers. We also show the matches derived from each step in the matching algorithm.

receive orders up to three months in advance, and the planning department may include such orders in the earliest planning report after receiving them rather than including them in the planning report for the month in which they will be completed. On the other, when dealing with repeated orders, the factory may start work early and then retrospectively include the order in a later planning report. For these reasons, we also include the three months of planning data preceding and following each year being matched.
The matching algorithm follows a 5-step process, where matched styles at each step are removed from the set of styles that are carried on to the next step.

1. We first look for direct matches between the style identifier in production and the following information in the planning data (in ranked order): style identifier, order number, purchase orders, and order IDs.

2. Next we attempt to match the style identifier in production to unmodified sub-strings in the same four variables in the planning data (style identifier, order number, purchase orders, and order IDs), in ranked order.

3. Then, we remove space and special characters from the following five variables in the planning data: order number, order IDs, order description, purchase orders, and style identifier. Having removed these characters, we attempt to match the style identifier in production to the modified variables, following the provided ranked order.

4. At this stage, we remove all spaces and special characters from the style identifiers in production and attempt to find direct matches in the versions of the style identifiers and order numbers from planning that were modified in the previous step.

5. Finally, we take the first five characters of the modified style identifiers in production and attempt to match them to sub-strings in the modified purchase orders from planning.

Table 3..3 shows the match rates in terms of observations in production when we take this library to the production data. Specifically it shows the number of matches, non-matches, and the percentage of non-matches by year, by unit, and overall. As shown at the bottom of the table, we successfully link styles between planning and production for roughly 52% of all production data. Table 3..2 shows that the mean differences in our key variables between the matched and non-matched observations are rather small. This suggests that, even if there was anything systematic affecting the match rates, this is not related to factors that in turn affect production outcomes and their determinants.

TABLE 3..2. DIFFERENCES BETWEEN MATCHED AND NON-MATCHED OBSERVATIONS

	Mean		Difference in means
	Matched	Not matched	$\frac{Matched-Notmatched}{Matched}$
Production data			
Efficiency	0.42	0.44	-0.03
Alteration rate	0.07	0.08	-0.13
Rejection rate	0.01	0.01	0.02
Style SMV	6.80	6.32	0.07
SD of SMV (line/day level)	0.51	0.46	0.10
Number of styles produced each day by line	1.79	1.81	-0.01
Running days - Reset at 7	5.37	5.15	0.04
Days passed	1.38	1.36	0.01
Daily output	915.67	998.57	-0.09

Notes: Table 3..2 shows the average differences in key variables in the production data between the observations that were successfully linked to the planning data and those that were not.

TABLE 3..3. MATCH RATES BY YEAR AND UNIT

	Matched	Not matched	% Not matched
By year			
2015	10309	13443	56.60
2016	16924	12030	41.55
2017	21373	16684	43.84
2018	32124	11515	26.39
2019	16980	29693	63.62
2020	2504	8975	78.19
By unit			
Unit 1	13455	10961	44.89
Unit 2	12215	10809	46.95
Unit 3	10341	7732	42.78
Unit 4	9233	5723	38.27
Unit 5	7446	8924	54.51
Unit 6	9975	8785	46.83
Unit 7	7587	10088	57.07
Unit 8	8291	7527	47.59
Unit 9	9167	7652	45.50
Unit 10	8868	6674	42.94
Unit 11	2546	5040	66.44
Overall	100214	92340	47.96

Notes: Table 3..3 displays the number of observations in the production data which could be matched and not matched 1-to-1 with a corresponding style in the planning data by year and by unit. It also shows the number of non-matched observations in percentage terms.

3.A Additional figures

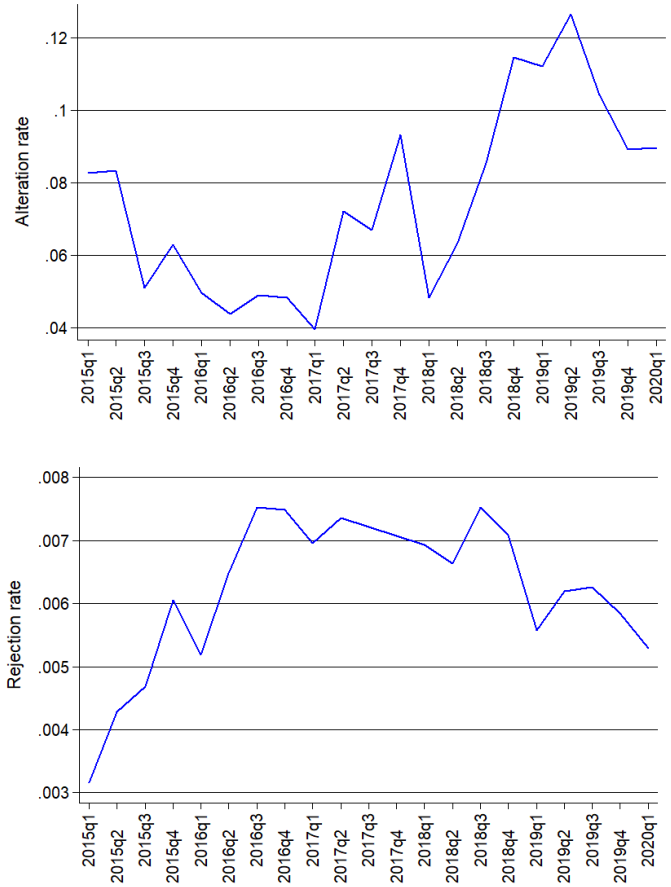


Figure 3.A.1: Alteration rates and rejection rates over time

Notes: The upper graph of figure 3.A.1 shows the evolution of average alteration rates in the study factory across time, whereas the lower graph shows it for rejection rates. Alteration rates are defined as the percentage of pieces per day that require additional work to fix a defect. Rejection rates are measured by the proportion of items per day rejected at the end of the line.

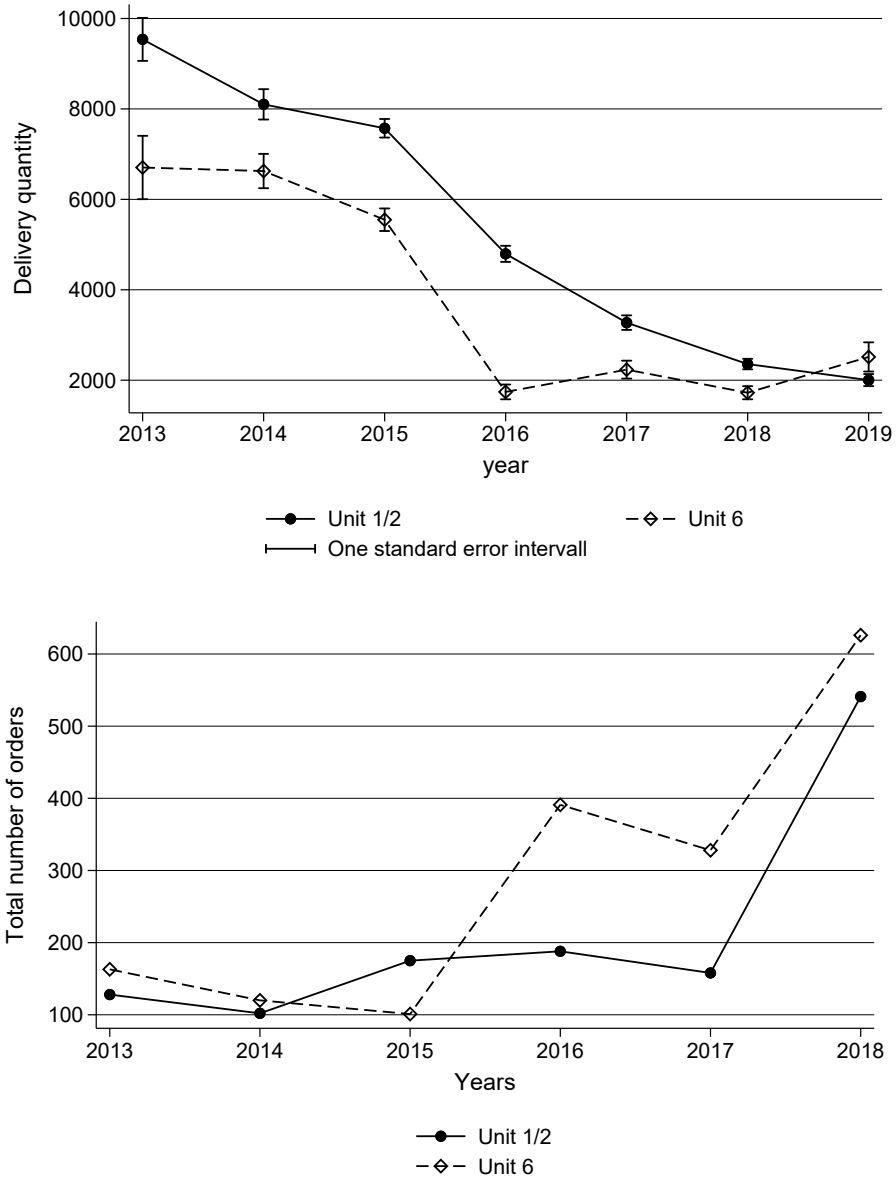


Figure 3.A.2: Imposed industry demand by unit

Notes: The upper graph of Figure 3.A.2 shows the median and standard deviations of delivery quantity across years for units 1 and 2, and unit 6. Delivery quantity is calculated as the sum of daily planned quantities per style and delivery date. The lower graph shows the total number of orders across years in units 1 and 2, and unit 6. In 2019, 79 percent of the order IDs are missing, which implies that the total order size is only based on a small share of all orders in that year. Hence, we do not show the total number of orders in 2019.

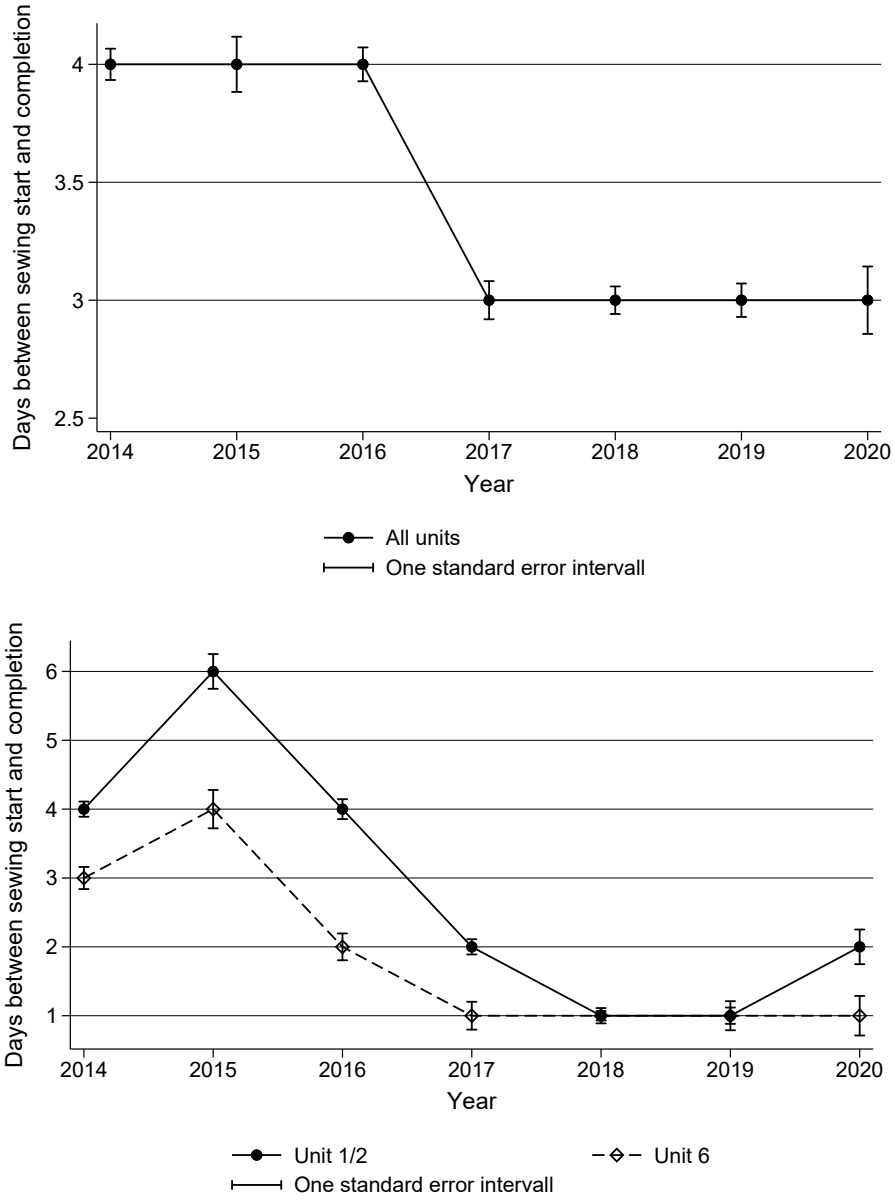


Figure 3.A.3: Days between sewing start and sewing end date

Notes: The upper graph of Figure 3.A.3 shows the number of days between sewing start date and sewing end date in the whole factory, whereas the lower graph shows it for units one/two and unit six.

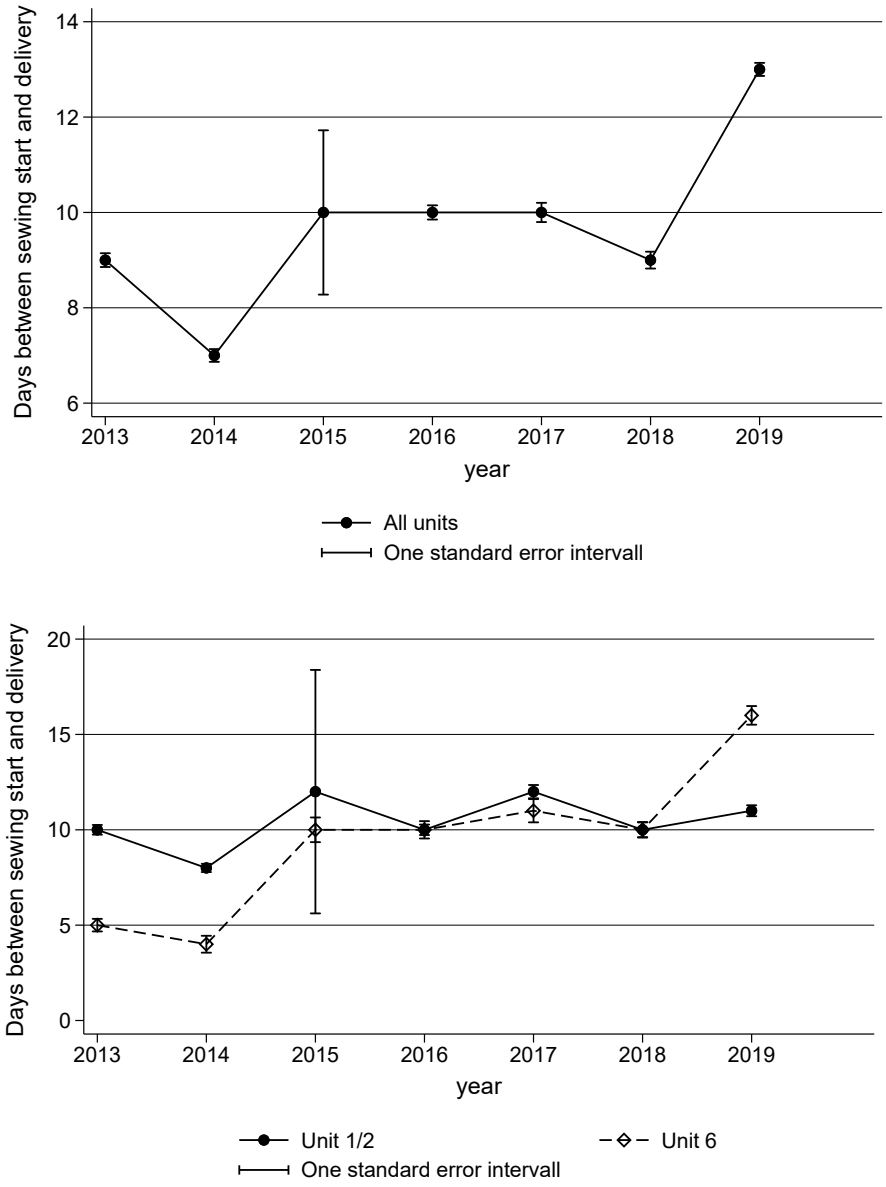


Figure 3.A.4: Days between sewing start and delivery date

Notes: The upper graph of Figure 3.A.4 shows the number of days between sewing start date and the delivery date in the whole factory, whereas the lower graph shows it in units one/two and unit six.

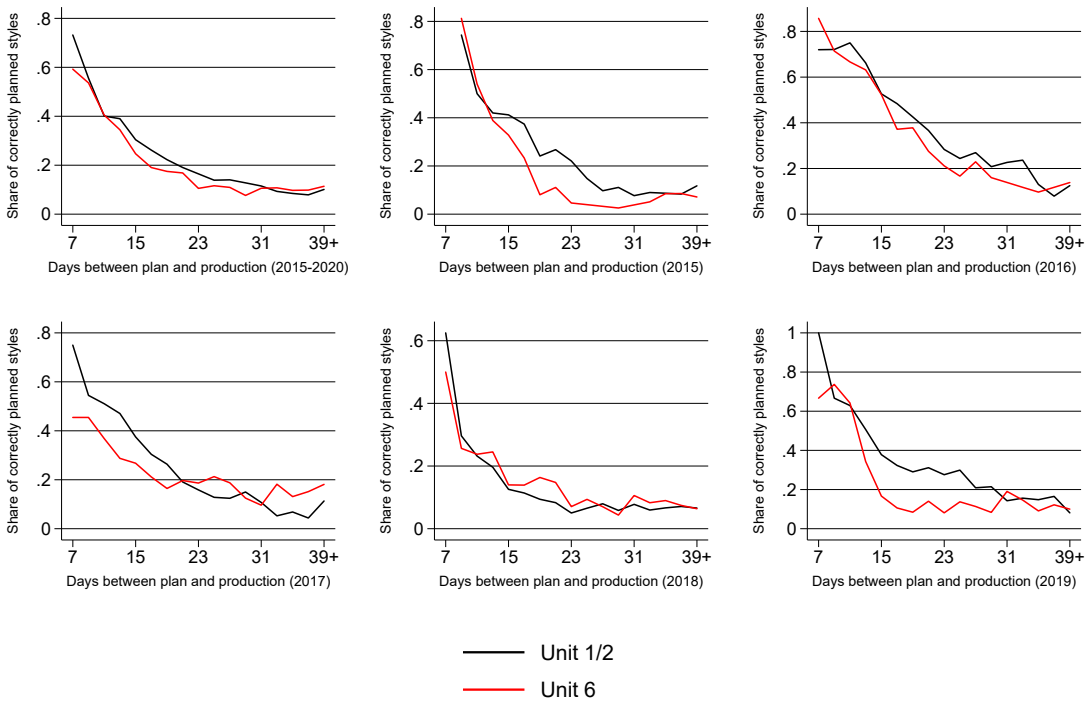


Figure 3.A.5: Percentage of correctly planned styles by unit

Notes: Figure 3.A.5 shows the share of "correctly" planned styles produced on the planned line and day by planning horizon across years. Units 1 and 2 are marked by a black line, and Unit 6 by a red line.

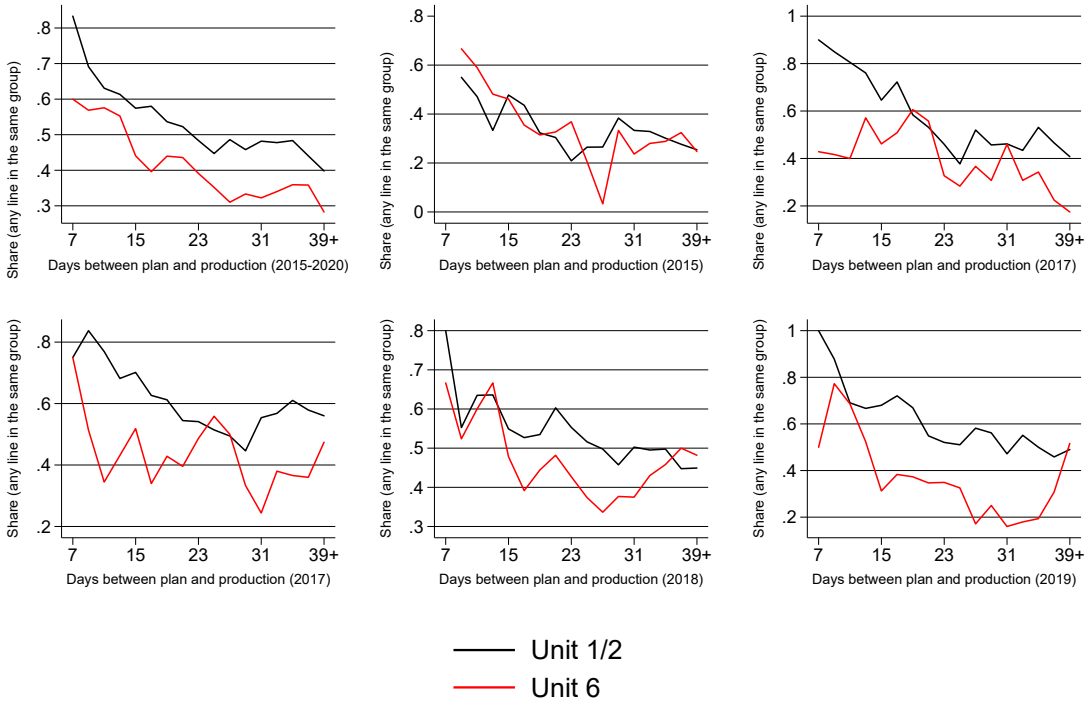


Figure 3.A.6: Percentage of correctly planned styles on any line (within the same group) by unit

Notes: Figure 3.A.6 shows the percentage of "correctly" planned styles on *any* production line within the same organisational group at a given day by planning horizon across years in Units 1 and 2, marked by a black line, and Unit 6, marked by a red line. The factory is composed of three organisational groups with three to four units per group.

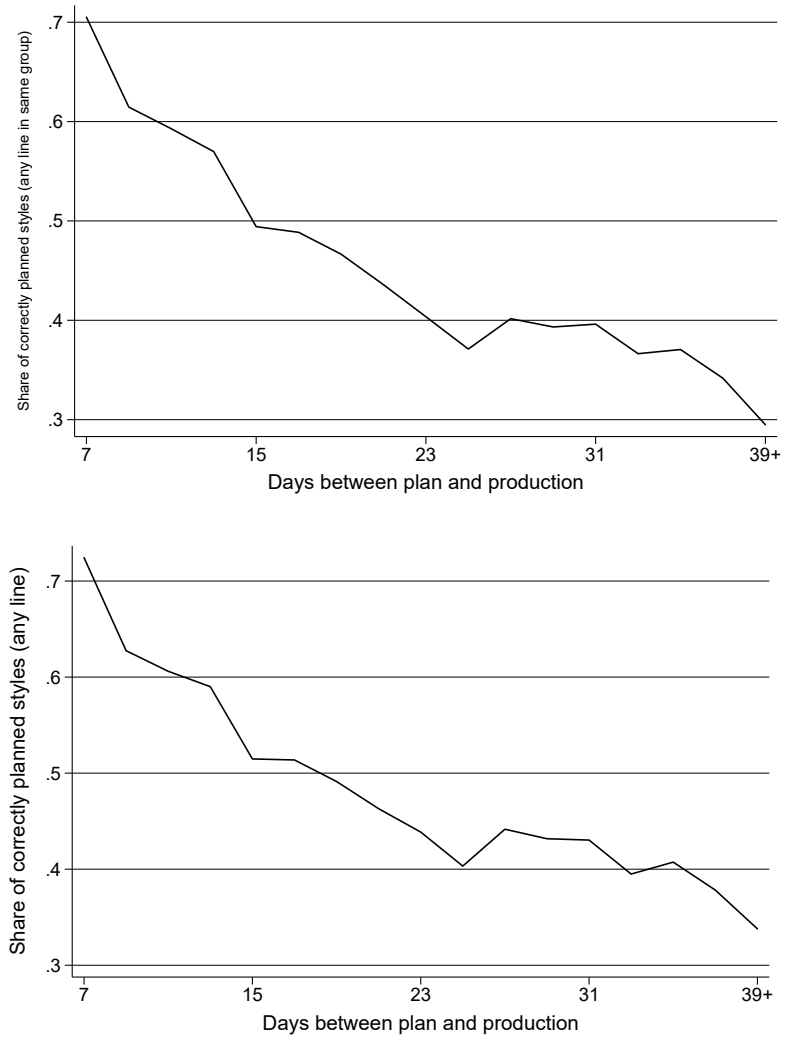


Figure 3.A.7: Percentage of correctly planned styles on any line within the same group and across the whole factory by planning horizon

Notes: The upper graph of figure 3.A.7 shows the percentage of correctly planned styles on *any* production line within the same organisational group at a given day by planning horizon. The factory is composed of three organisational groups with three to four units per group. The lower graph shows the percentage of correctly planned styles on *any* production line within the whole factory at a given day by planning horizon. The planning horizon is the number of days between the production date and the week the plan was made. As we take the first day of the respective week, the x-axis starts at 7 days.

3.B Additional tables

TABLE 3.B.4. OVERVIEW OF VARIABLES RELATED TO PRODUCTION

		Unit 1	Unit 2	Unit 3	Unit 4	Unit 5	Unit 6	Unit 7	Unit 8	Unit 9
Efficiency	**2015**	**55%**	**54%**	**50%**	**41%**	**51%**	**47%**	**56%**	**55%**	**43%**
	delta 2015-2017	*-29%*	*-22%*	*-29%*	*-28%*	*-43%*	*-45%*	*-51%*	*-33%*	*-15%*
	delta 2015-2019	*-13%*	*-14%*	*-6%*	*-6%*	*-6%*	*-29%*	*-18%*	*-12%*	*11%*
Alteration rates	**2015**	**6%**	**8%**	**7%**	**7%**	**8%**	**8%**	**7%**	**8%**	**6%**
	delta 2015-2017	*-3%*	*-37%*	*-11%*	*-9%*	*-19%*	*-4%*	*19%*	*-2%*	*23%*
	delta 2015-2019	*36%*	*-5%*	*14%*	*64%*	*61%*	*131%*	*54%*	*24%*	*62%*
Reject rates	**2015**	**0,4%**	**0,4%**	**0,2%**	**0,5%**	**0,7%**	**0,5%**	**0,4%**	**0,4%**	**0,4%**
	delta 2015-2017	*132%*	*209%*	*279%*	*-16%*	*11%*	*28%*	*215%*	*-48%*	*-50%*
	delta 2015-2019	*45%*	*90%*	*170%*	*7%*	*-11%*	*58%*	*83%*	*21%*	*-27%*
Running days (capped at 7)	**2015**	**5,89**	**5,40**	**5,44**	**5,14**	**5,09**	**4,83**	**5,43**	**5,33**	**5,69**
	delta 2015-2017	*8%*	*6%*	*14%*	*4%*	*-16%*	*5%*	*7%*	*1%*	*8%*
	delta 2015-2019	*-22%*	*-12%*	*-15%*	*-19%*	*-5%*	*-2%*	*-13%*	*11%*	*5%*
Days passed	**2015**	**1,14**	**1,17**	**1,15**	**1,07**	**1,07**	**1,07**	**1,09**	**1,09**	**1,19**
	delta 2015-2017	*18%*	*14%*	*5%*	*13%*	*50%*	*29%*	*23%*	*8%*	*-8%*
	delta 2015-2019	*59%*	*36%*	*48%*	*135%*	*18%*	*43%*	*40%*	*5%*	*2%*
Style SMV	**2015**	**6,89**	**7,14**	**8,19**	**9,35**	**7,50**	**7,43**	**7,49**	**6,29**	**4,55**
	delta 2015-2017	*-8%*	*-18%*	*-12%*	*-1%*	*-13%*	*-16%*	*-35%*	*14%*	*60%*
	delta 2015-2019	*-27%*	*-27%*	*-27%*	*-22%*	*-33%*	*-11%*	*-33%*	*4%*	*48%*
SD SMV (line level)	**2015**	**0,33**	**0,40**	**0,41**	**0,42**	**0,36**	**0,37**	**0,32**	**0,26**	**0,07**
	delta 2015-2017	*46%*	*-6%*	*24%*	*70%*	*11%*	*15%*	*79%*	*34%*	*423%*
	delta 2015-2019	*33%*	*-9%*	*61%*	*67%*	*-41%*	*42%*	*0%*	*69%*	*654%*
Number of styles produced each day by line	**2015**	**1,56**	**1,62**	**1,54**	**1,37**	**1,43**	**1,50**	**1,51**	**1,48**	**1,52**
	delta 2015-2017	*26%*	*14%*	*13%*	*7%*	*9%*	*31%*	*28%*	*7%*	*-1%*
	delta 2015-2019	*41%*	*34%*	*39%*	*21%*	*29%*	*34%*	*25%*	*7%*	*3%*
Proportional change in SMV day average	**2015**	**0,05**	**0,06**	**0,06**	**0,07**	**0,06**	**0,06**	**0,06**	**0,04**	**0,02**
	delta 2015-2017	*34%*	*5%*	*22%*	*42%*	*46%*	*24%*	*13%*	*37%*	*170%*
	delta 2015-2019	*56%*	*11%*	*50%*	*46%*	*-30%*	*41%*	*-2%*	*46%*	*234%*

Notes: Table 3.B.4 displays the averages of key variables in the production data in 2015 per unit. Below the 2015 average for each variable the table also shows the percentage change between 2017 and 2015, and 2019 and 2015.

TABLE 3.B.5. STANDARD SPECIFICATION WITH LINE FIXED EFFECTS

	Efficiency	Alteration Rate	Rejection Rate
	(1)	(2)	(3)
Style SMV	-0.676***	0.327***	0.018***
	(0.000)	(0.000)	(0.000)
First day style is produced on line X	-2.438***	0.004	0.006
	(0.000)	(0.945)	(0.635)
SD of SMV (line level)	-0.846***	0.055***	0.023***
	(0.000)	(0.000)	(0.000)
Number of styles produced each day by line	-1.809***	-0.043***	0.019***
	(0.000)	(0.000)	(0.000)
Proportional change in SMV day average	-15.496***	1.210***	0.019
	(0.000)	(0.000)	(0.416)
Cumulative running days - Line	0.077***	-0.024***	0.000
	(0.000)	(0.000)	(0.728)
Cumulative running days - Unit	0.009*	0.005***	0.001***
	(0.088)	(0.000)	(0.000)
Running days (capped at 7)=1	-7.198***	0.430***	0.010
	(0.000)	(0.000)	(0.504)
Running days (capped at 7)=2	-2.818***	0.349***	-0.026***
	(0.000)	(0.000)	(0.003)
Running days (capped at 7)=3	-1.219***	0.145***	-0.020**
	(0.000)	(0.000)	(0.028)
Running days (capped at 7)=4	-0.321*	0.028	-0.028***
	(0.060)	(0.431)	(0.001)
Running days (capped at 7)=5	-0.333*	0.006	-0.019**
	(0.057)	(0.876)	(0.027)
Running days (capped at 7)=6	-0.434**	0.028	-0.008
	(0.018)	(0.460)	(0.411)
Running days (capped at 7)=7	0.000	0.000	0.000
	(.)	(.)	(.)
Capped Days (6) passed=1	0.000	0.000	0.000
	(.)	(.)	(.)
Capped Days (6) passed=2	-5.768***	0.002	-0.007
	(0.000)	(0.956)	(0.522)
Capped Days (6) passed=3	-7.831***	-0.080	0.001
	(0.000)	(0.189)	(0.946)
Capped Days (6) passed=4	-8.464***	-0.100	-0.016
	(0.000)	(0.192)	(0.382)
Capped Days (6) passed=5	-8.225***	0.016	-0.013
	(0.000)	(0.865)	(0.612)
Capped Days (6) passed=6	-4.829***	-0.329***	-0.011
	(0.000)	(0.000)	(0.490)
Constant	48.954***	7.336***	0.094
	(0.000)	(0.000)	(0.190)
RD (line)=RD (unit)	0.00	0.00	0.43
Line FE	Yes	Yes	Yes
Month/year FE	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes

Notes: Running days are reset at days passed=7, capped at 7+, and have 7+ as the base category. Days passed is capped at 6+, with 1 as the base category). We use robust standard errors and allow for line level, month/year, and buyer fixed effects. * p<0.1, ** p<0.05, *** p<0.01

TABLE 3.B.6. STANDARD SPECIFICATION WITH LINE FE AND CLUSTERED SE

	Efficiency	Alteration Rate	Rejection Rate
	(1)	(2)	(3)
Style SMV	-0.676***	0.327***	0.018***
	(0.000)	(0.000)	(0.000)
First day style is produced on line X	-2.438***	0.004	0.006
	(0.000)	(0.947)	(0.658)
SD of SMV (line level)	-0.846***	0.055***	0.023***
	(0.000)	(0.007)	(0.000)
Number of styles produced each day by line	-1.809***	-0.043**	0.019***
	(0.000)	(0.014)	(0.000)
Proportional change in SMV day average	-15.496***	1.210***	0.019
	(0.000)	(0.000)	(0.444)
Cumulative running days - Line	0.077**	-0.024***	0.000
	(0.033)	(0.001)	(0.853)
Cumulative running days - Unit	0.009	0.005*	0.001*
	(0.451)	(0.092)	(0.055)
Running days (capped at 7)=1	-7.198***	0.430***	0.010
	(0.000)	(0.000)	(0.581)
Running days (capped at 7)=2	-2.818***	0.349***	-0.026**
	(0.000)	(0.000)	(0.043)
Running days (capped at 7)=3	-1.219***	0.145***	-0.020
	(0.000)	(0.006)	(0.102)
Running days (capped at 7)=4	-0.321	0.028	-0.028**
	(0.206)	(0.565)	(0.014)
Running days (capped at 7)=5	-0.333	0.006	-0.019*
	(0.156)	(0.897)	(0.074)
Running days (capped at 7)=6	-0.434**	0.028	-0.008
	(0.038)	(0.465)	(0.445)
Running days (capped at 7)=7	0.000	0.000	0.000
	(.)	(.)	(.)
Capped Days (6) passed=1	0.000	0.000	0.000
	(.)	(.)	(.)
Capped Days (6) passed=2	-5.768***	0.002	-0.007
	(0.000)	(0.959)	(0.537)
Capped Days (6) passed=3	-7.831***	-0.080	0.001
	(0.000)	(0.184)	(0.947)
Capped Days (6) passed=4	-8.464***	-0.100	-0.016
	(0.000)	(0.202)	(0.385)
Capped Days (6) passed=5	-8.225***	0.016	-0.013
	(0.000)	(0.867)	(0.622)
Capped Days (6) passed=6	-4.829***	-0.329***	-0.011
	(0.000)	(0.000)	(0.527)
Constant	48.954***	7.336***	0.094
	(0.000)	(0.000)	(0.227)
RD (line)=RD (unit)	0.11	0.00	0.67
Line FE	Yes	Yes	Yes
Month/year FE	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes

Notes: Running days are reset at days passed=7, capped at 7+, and have 7+ as the base category. Days passed is capped at 6+, with 1 as the base category). We cluster robust standard errors on style identifiers and allow for line level, month/year, and buyer fixed effects. * p<0.1, ** p<0.05, *** p<0.01

TABLE 3.B.7. STANDARD SPECIFICATION BY YEAR

	Full period	2015	2016	2017	2018	2019	2020
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Style SMV	-0.852***	-0.934***	-0.927***	-0.750***	-0.981***	-0.700***	-0.976***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
First day style is produced on line X	-2.506***	-4.919***	-1.254*	-2.192***	-3.063***	-2.063***	-1.330
	(0.000)	(0.000)	(0.097)	(0.000)	(0.000)	(0.000)	(0.648)
SD of SMV (line level)	-0.722***	1.197	-1.043	-0.295***	-0.839***	-1.090***	-1.649***
	(0.000)	(0.296)	(0.146)	(0.008)	(0.000)	(0.000)	(0.000)
Number of styles produced each day by line	-1.928***	3.847***	3.485***	-1.499***	-2.541***	-1.708***	-3.166***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Proportional change in SMV day average	-15.259***	-20.327***	-13.192***	-12.780***	-11.591***	-15.406***	-21.703***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Cumulative running days - Line	0.135***	0.267***	0.209***	0.222***	0.096***	0.068*	0.044
	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)	(0.051)	(0.552)
Cumulative running days - Unit	-0.009*	0.039	-0.137***	-0.039***	-0.007	-0.003	0.123***
	(0.072)	(0.103)	(0.000)	(0.000)	(0.536)	(0.745)	(0.000)
Running days (capped at 7)=1	-7.446***	-7.293***	-12.842***	-7.192***	-7.660***	-6.952***	-4.145
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.149)
Running days (capped at 7)=2	-3.101***	-5.534***	-7.967***	-2.600***	-3.126***	-1.104***	-1.586**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.003)	(0.038)
Running days (capped at 7)=3	-1.508***	-2.623***	-4.565***	-0.658*	-2.082***	-0.251	-0.347
	(0.000)	(0.000)	(0.000)	(0.056)	(0.000)	(0.492)	(0.631)
Running days (capped at 7)=4	-0.578***	-1.435**	-2.428***	0.229	-1.043***	0.141	0.067
	(0.001)	(0.012)	(0.000)	(0.513)	(0.005)	(0.701)	(0.924)
Running days (capped at 7)=5	-0.547***	-0.496	-1.733***	0.194	-0.854**	-0.425	-0.331
	(0.003)	(0.378)	(0.001)	(0.592)	(0.028)	(0.252)	(0.634)
Running days (capped at 7)=6	-0.630***	-0.929	-1.465***	-0.146	-1.147***	-0.495	0.951
	(0.001)	(0.110)	(0.007)	(0.702)	(0.005)	(0.196)	(0.178)
Running days (capped at 7)=7	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)	(.)
Capped Days (6) passed=1	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)	(.)
Capped Days (6) passed=2	-6.043***	-1.904**	0.667	-5.159***	-7.683***	-7.922***	-3.790***
	(0.000)	(0.012)	(0.362)	(0.000)	(0.000)	(0.000)	(0.000)
Capped Days (6) passed=3	-7.965***	-3.281	0.633	-5.657***	-7.889***	-10.773***	-7.902***
	(0.000)	(0.105)	(0.714)	(0.000)	(0.000)	(0.000)	(0.000)
Capped Days (6) passed=4	-8.931***	-3.007	-2.756	-6.886***	-10.656***	-9.430***	-6.512***
	(0.000)	(0.257)	(0.237)	(0.000)	(0.000)	(0.000)	(0.000)
Capped Days (6) passed=5	-8.810***	-2.552	-5.883	-5.731***	-9.645***	-11.402***	-4.563**
	(0.000)	(0.416)	(0.415)	(0.000)	(0.000)	(0.000)	(0.049)
Capped Days (6) passed=6	-4.907***	-3.968***	-2.892*	-3.543***	-5.250***	-5.550***	-2.686
	(0.000)	(0.003)	(0.062)	(0.000)	(0.000)	(0.000)	(0.291)
Constant	38.756***	31.686***	65.749***	48.445***	62.699***	57.860***	63.959***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
RD (line)=RD (unit)	0.00	0.00	0.00	0.00	0.01	0.07	0.35
Line FE	No	No	No	No	No	No	No
Month/year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Running days are reset at days passed=7, capped at 7+, and have 7+ as the base category. Days passed is capped at 6+, with 1 as the base category). We use robust standard errors and allow for month/year, and buyer fixed effects. * p<0.1, ** p<0.05, *** p<0.01

TABLE 3.B.8. SELECTED UNITS WITH STYLE FIXED EFFECTS

	Efficiency	Alteration Rate	Rejection Rate
	(1)	(2)	(3)
Unit=1	7.683***	-0.783***	-0.319***
	(0.000)	(0.000)	(0.000)
Unit=2	6.891***	-0.444***	-0.252***
	(0.000)	(0.001)	(0.000)
Unit=6	0.000	0.000	0.000
	(.)	(.)	(.)
Style SMV	-16.392***	1.919**	1.724***
	(0.001)	(0.036)	(0.000)
SD of SMV (line level)	-0.849	0.217***	0.026
	(0.304)	(0.006)	(0.231)
Proportional change in SMV day average	-18.142***	2.182***	0.196
	(0.000)	(0.000)	(0.189)
Number of styles produced each day by line	-1.151**	-0.059	-0.001
	(0.043)	(0.409)	(0.953)
Cumulative running days - Line	0.183	0.041*	0.026***
	(0.115)	(0.073)	(0.005)
Cumulative running days - Unit	-0.005	-0.009	-0.004
	(0.912)	(0.277)	(0.235)
First day style is produced on line X	-2.446	-0.193	-0.056
	(0.212)	(0.581)	(0.482)
Running days (capped at 7)=1	-7.881***	0.969**	0.166
	(0.001)	(0.016)	(0.117)
Running days (capped at 7)=2	-6.227***	0.796***	0.080
	(0.000)	(0.001)	(0.253)
Running days (capped at 7)=3	-3.142***	0.333	0.076
	(0.004)	(0.101)	(0.232)
Running days (capped at 7)=4	-2.594**	0.453**	0.089
	(0.017)	(0.018)	(0.168)
Running days (capped at 7)=5	-1.506	0.190	0.109*
	(0.173)	(0.300)	(0.063)
Running days (capped at 7)=6	-2.146**	0.372*	0.074
	(0.040)	(0.052)	(0.200)
Running days (capped at 7)=7	0.000	0.000	0.000
	(.)	(.)	(.)
Capped Days (6) passed=1	0.000	0.000	0.000
	(.)	(.)	(.)
Capped Days (6) passed=2	-3.678**	0.082	-0.122***
	(0.012)	(0.705)	(0.006)
Capped Days (6) passed=3	-11.070***	0.320	0.088
	(0.002)	(0.383)	(0.516)
Capped Days (6) passed=4	-8.144**	-0.248	0.066
	(0.023)	(0.540)	(0.681)
Capped Days (6) passed=5	-0.451	-0.122	0.006
	(0.950)	(0.824)	(0.960)
Capped Days (6) passed=6	-4.872*	-0.460	-0.263***
	(0.051)	(0.224)	(0.006)
Constant	164.661***	-6.026	-10.603***
	(0.000)	(0.305)	(0.000)
RD (line)=RD (unit)	0.19	0.07	0.01
Line FE	No	No	No
Month/year FE	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes

Notes: Running days are reset at days passed=7, capped at 7+, and have 7+ as the base category. Days passed is capped at 6+, with 1 as the base category). We use robust standard errors and allow for style, unit, month/year, and buyer fixed effects. * p<0.1, ** p<0.05, *** p<0.01

TABLE 3.B.9. EFFECTS ON EFFICIENCY USING ALL STYLES AND COMMON STYLES IN SELECTED UNITS (FULL TABLE)

	All styles (Full period)	Common styles (Full period)	All styles (2015)	Common styles (2015)	All styles (2017)	Common styles (2017)	All styles (2019)	Common styles (2019)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Unit=1	15.145***	7.683***	8.073***	5.537***	12.363***	9.267***	15.769***	2.368
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.793)
Unit=2	15.157***	6.891***	7.732***	5.519***	14.644***	7.553***	14.208***	12.020*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.085)
Unit=6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)
Style SMV	-1.075***	-16.392***	-1.038***	15.129	-1.029***	-21.565	-0.626***	-63.187
	(0.000)	(0.001)	(0.000)	(0.108)	(0.000)	(0.546)	(0.000)	(0.122)
First day style is produced on line X	-2.792***	-2.446	-5.285***	-4.482	-3.003***	1.230	-2.105**	0.512
	(0.000)	(0.212)	(0.000)	(0.155)	(0.005)	(0.784)	(0.026)	(0.890)
SD of SMV (line level)	-1.751***	-0.849	-0.233	4.362	-1.516***	0.466	-1.794***	-5.783
	(0.000)	(0.304)	(0.919)	(0.169)	(0.000)	(0.731)	(0.000)	(0.160)
Number of styles produced each day by line	-2.226***	-1.151**	3.234***	2.396**	-1.345***	-1.114	-2.896***	-3.029
	(0.000)	(0.043)	(0.000)	(0.018)	(0.000)	(0.275)	(0.000)	(0.242)
Proportional change in SMV day average	-14.246***	-18.142***	-22.798***	-24.424***	-12.042***	-8.165	-10.922***	-1.960
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.251)	(0.000)	(0.928)
Cumulative running days - Line	0.069**	0.183	0.413***	0.808***	0.183***	-0.473	-0.092	0.554
	(0.016)	(0.115)	(0.000)	(0.000)	(0.000)	(0.168)	(0.143)	(0.335)
Cumulative running days - Unit	0.012	-0.005	-0.053	-0.029	-0.011	0.288**	0.068***	-0.212
	(0.183)	(0.912)	(0.207)	(0.779)	(0.411)	(0.046)	(0.000)	(0.608)
Running days (capped at 7)=1	-5.246***	-7.881***	-3.993**	-0.479	-6.079***	-14.765**	-4.122***	-8.174
	(0.000)	(0.001)	(0.034)	(0.903)	(0.000)	(0.012)	(0.000)	(0.151)
Running days (capped at 7)=2	-2.522***	-6.227***	-4.617***	-2.138	-1.904***	-6.441*	-0.482	3.511
	(0.000)	(0.000)	(0.000)	(0.202)	(0.001)	(0.058)	(0.452)	(0.412)
Running days (capped at 7)=3	-1.300***	-3.142***	-2.323***	-0.164	-0.462	-3.346	-0.047	5.984
	(0.000)	(0.004)	(0.009)	(0.914)	(0.431)	(0.304)	(0.940)	(0.153)
Running days (capped at 7)=4	-0.501	-2.594**	-1.702**	0.539	0.760	-1.506	0.286	1.674
	(0.102)	(0.017)	(0.047)	(0.720)	(0.200)	(0.615)	(0.657)	(0.753)
Running days (capped at 7)=5	-0.336	-1.506	-0.353	1.732	0.375	-2.873	-0.292	-0.815
	(0.283)	(0.173)	(0.680)	(0.237)	(0.541)	(0.362)	(0.657)	(0.903)
Running days (capped at 7)=6	-0.626*	-2.146**	-0.199	0.875	0.391	-5.399*	-0.636	4.037
	(0.058)	(0.040)	(0.818)	(0.539)	(0.561)	(0.074)	(0.361)	(0.432)
Running days (capped at 7)=7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)
Capped Days (6) passed=1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)
Capped Days (6) passed=2	-6.344***	-3.678**	-3.187***	-1.134	-5.518***	-4.352	-7.660***	-11.569
	(0.000)	(0.012)	(0.004)	(0.510)	(0.000)	(0.222)	(0.000)	(0.343)
Capped Days (6) passed=3	-7.180***	-11.070***	-1.603	4.709	-6.346***	-12.615***	-9.251***	0.877
	(0.000)	(0.002)	(0.568)	(0.239)	(0.000)	(0.001)	(0.000)	(0.864)
Capped Days (6) passed=4	-8.724***	-8.144**	-2.205	-3.361	-8.080***	-16.920***	-8.239***	3.913
	(0.000)	(0.023)	(0.613)	(0.506)	(0.000)	(0.002)	(0.000)	(0.597)
Capped Days (6) passed=5	-8.140***	-0.451	-2.226	15.578*	-7.185***	-14.542***	-8.978***	-8.932
	(0.000)	(0.950)	(0.737)	(0.096)	(0.000)	(0.000)	(0.000)	(0.294)
Capped Days (6) passed=6	-6.826***	-4.872*	-4.573**	-12.708***	-5.677***	-2.715	-7.101***	4.671
	(0.000)	(0.051)	(0.023)	(0.001)	(0.000)	(0.654)	(0.000)	(0.448)
Constant	32.332***	164.661***	20.418***	-71.807	39.344***	168.654	46.394***	368.019
	(0.000)	(0.000)	(0.000)	(0.261)	(0.000)	(0.396)	(0.000)	(0.109)
Unit 1=Unit 2	0.94	0.19	0.46	0.98	0.00	0.34	0.00	0.16
Line FE	No	No	No	No	No	No	No	No
Month/year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Style FE	No	Yes	No	Yes	No	Yes	No	Yes

Notes: Running days are reset at days passed=7, capped at 7+, and have 7+ as the base category. Days passed is capped at 6+, with 1 as the base category). We use robust standard errors and allow month/year, buyer, and unit (with unit 6 as the base category) fixed effects in all columns. Even columns only use style in common across the three units and also include style fixed effects.
* p<0.1, ** p<0.05, *** p<0.01"

TABLE 3.B.10. STANDARD SPECIFICATION FOR INDUSTRY

	Efficiency	Alteration Rate	Rejection Rate
	(1)	(2)	(3)
Style SMV	-0.833***	0.318***	0.028***
	(0.000)	(0.000)	(0.000)
First day style is produced on line X	-2.140***	0.367***	0.030*
	(0.000)	(0.000)	(0.098)
SD of SMV (line level)	-2.534***	0.281***	0.028***
	(0.000)	(0.000)	(0.000)
Number of styles produced each day by line	-3.378***	0.084***	-0.004
	(0.000)	(0.000)	(0.148)
Proportional change in SMV day average	-27.105***	3.764***	0.238***
	(0.000)	(0.000)	(0.000)
Cumulative running days - Line	0.114***	-0.024***	-0.001***
	(0.000)	(0.000)	(0.000)
Cumulative running days - Unit	-0.027***	0.001**	0.000***
	(0.000)	(0.041)	(0.000)
Running days (capped at 7)=1	-4.857***	0.174*	0.014
	(0.000)	(0.094)	(0.467)
Running days (capped at 7)=2	-2.968***	0.355***	0.012*
	(0.000)	(0.000)	(0.069)
Running days (capped at 7)=3	-1.839***	0.232***	0.003
	(0.000)	(0.000)	(0.536)
Running days (capped at 7)=4	-1.079***	0.111**	-0.001
	(0.000)	(0.012)	(0.924)
Running days (capped at 7)=5	-0.419**	0.072	-0.006
	(0.018)	(0.128)	(0.436)
Running days (capped at 7)=6	-0.124	0.073	-0.017***
	(0.507)	(0.152)	(0.004)
Running days (capped at 7)=7	0.000	0.000	0.000
	(.)	(.)	(.)
Capped Days (6) passed=1	0.000	0.000	0.000
	(.)	(.)	(.)
Capped Days (6) passed=2	0.693**	-0.200***	-0.025***
	(0.011)	(0.005)	(0.002)
Capped Days (6) passed=3	0.837*	-0.325***	0.001
	(0.079)	(0.001)	(0.934)
Capped Days (6) passed=4	-0.196	-0.056	-0.002
	(0.777)	(0.707)	(0.917)
Capped Days (6) passed=5	0.836	-0.090	-0.007
	(0.368)	(0.627)	(0.682)
Capped Days (6) passed=6	2.126***	-0.189*	-0.001
	(0.000)	(0.099)	(0.968)
Constant	55.213***	4.454***	0.066
	(0.000)	(0.000)	(0.484)
RD (line)=RD (unit)	0.00	0.00	0.00
Line FE	No	No	No
Month/year FE	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes

Notes: Running days are reset at days passed=7, capped at 7+, and have 7+ as the base category. Days passed is capped at 6+, with 1 as the base category). We use robust standard errors and allow for month/year, and buyer fixed effects. * p<0.1, ** p<0.05, *** p<0.01

TABLE 3.B.11. STANDARD SPECIFICATION BY UNIT

	Unit 1	Unit 2	Unit 1+2	Unit 3	Unit 4	Unit 5	Unit 6	Unit 7	Unit 8	Unit 9	Unit 10
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Style SMV	-1.424***	-1.407***	-1.438***	-0.589***	-0.636***	-1.046***	-0.028	-0.756***	-0.927***	-1.050***	-0.749***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.591)	(0.000)	(0.000)	(0.000)	(0.000)
First day style is produced on line X	-2.917***	-2.838***	-2.949***	-2.746***	-2.884***	-1.285	-3.131***	-0.490	-2.690***	-3.315***	-1.729*
	(0.000)	(0.001)	(0.000)	(0.003)	(0.001)	(0.183)	(0.000)	(0.593)	(0.006)	(0.000)	(0.057)
SD of SMV (line level)	-2.382***	-2.319***	-2.357***	-1.580***	-0.673***	-0.754***	-0.534***	-1.103***	-1.446***	-0.477**	-0.329
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.006)	(0.004)	(0.000)	(0.000)	(0.024)	(0.102)
Nb. styles produced per line-day	-2.410***	-2.306***	-2.361***	-0.664***	-1.422***	-1.546***	-1.443***	-1.916***	-1.528***	-2.334***	-2.395***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Prop. change in SMV day average	-14.945***	-15.003***	-14.852***	-8.168***	-10.028***	-24.206***	-13.535***	-19.129***	-15.930***	-13.355***	-13.870***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Cumulative running days - Line	0.021	0.051	0.045	0.132***	0.109*	0.020	-0.004	0.125**	0.142***	-0.025	-0.087
	(0.636)	(0.347)	(0.186)	(0.004)	(0.081)	(0.708)	(0.934)	(0.015)	(0.006)	(0.556)	(0.102)
Cumulative running days - Unit	0.067***	0.015	0.040***	0.081***	0.100***	-0.011	-0.033*	-0.068***	-0.063**	0.064***	0.076***
	(0.000)	(0.320)	(0.000)	(0.000)	(0.000)	(0.598)	(0.051)	(0.000)	(0.013)	(0.000)	(0.000)
Running days (capped at 7)=1	-5.084***	-7.131***	-5.968***	-6.376***	-8.577***	-5.534***	-2.696***	-4.794***	-10.259***	-11.292***	-12.761***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.003)	(0.000)	(0.000)	(0.000)	(0.000)
Running days (capped at 7)=2	-2.238***	-3.325***	-2.718***	-1.219**	-3.683***	-2.544***	-1.830***	0.208	-4.712***	-4.840***	-5.253***
	(0.000)	(0.000)	(0.000)	(0.024)	(0.000)	(0.000)	(0.000)	(0.709)	(0.000)	(0.000)	(0.000)
Running days (capped at 7)=3	-0.952*	-1.765***	-1.314***	-0.218	-1.440**	-1.910***	-1.102**	0.669	-1.490***	-1.305**	-1.540***
	(0.056)	(0.002)	(0.000)	(0.682)	(0.011)	(0.000)	(0.024)	(0.221)	(0.009)	(0.010)	(0.010)
Running days (capped at 7)=4	-0.103	-0.710	-0.371	1.440***	-0.162	-0.988*	-0.695	0.871	-0.230	-0.323	-0.704
	(0.843)	(0.190)	(0.320)	(0.006)	(0.768)	(0.074)	(0.162)	(0.114)	(0.692)	(0.528)	(0.226)
Running days (capped at 7)=5	0.351	-0.715	-0.118	0.739	-0.108	-1.033*	-0.960*	0.979*	-0.289	-0.184	-0.354
	(0.504)	(0.197)	(0.756)	(0.167)	(0.846)	(0.076)	(0.061)	(0.083)	(0.621)	(0.724)	(0.551)
Running days (capped at 7)=6	0.175	-1.368**	-0.545	-0.025	-0.392	-0.064	-0.913*	0.602	-0.947	-0.893*	-0.287
	(0.757)	(0.017)	(0.173)	(0.966)	(0.484)	(0.916)	(0.091)	(0.299)	(0.123)	(0.097)	(0.648)
Running days (capped at 7)=7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)
Capped Days (6) passed=1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)
Capped Days (6) passed=2	-7.496***	-5.630***	-6.610***	-5.529***	-2.186**	-6.646***	-6.069***	-5.327***	-7.075***	-4.965***	-4.559***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.011)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Capped Days (6) passed=3	-7.502***	-8.315***	-7.814***	-11.298***	-3.806**	-6.039***	-5.887***	-8.134***	-8.725***	-7.824***	-8.318***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.017)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Capped Days (6) passed=4	-10.119***	-10.471***	-10.205***	-10.633***	-5.924***	-5.878***	-5.434***	-8.560***	-10.773***	-11.656***	-5.676***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.003)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.006)
Capped Days (6) passed=5	-6.897***	-11.456***	-9.156***	-14.056***	-11.307***	-2.734	-5.306***	-5.902***	-10.614***	-7.262**	-5.679**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.243)	(0.000)	(0.001)	(0.000)	(0.016)	(0.048)
Capped Days (6) passed=6	-8.394***	-7.403***	-7.932***	-5.859***	-4.254***	-4.168***	-4.063***	-2.522**	-4.015***	-4.523***	2.528*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)	(0.000)	(0.044)	(0.005)	(0.002)	(0.084)
Constant	61.717***	66.135***	62.068***	54.771***	53.258***	58.120***	11.348***	54.380***	63.348***	62.818***	62.341***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
RD (line)=RD (unit)	0.37	0.57	0.92	0.40	0.90	0.66	0.63	0.00	0.00	0.10	0.02
Line FE	No	No	No	No	No	No	No	No	No	No	No
Month/year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Running days are reset at days passed=7, capped at 7+, and have 7+ as the base category. Days passed is capped at 6+, with 1 as the base category). We use robust standard errors and allow for month/year, and buyer fixed effects. * p<0.1, ** p<0.05, *** p<0.01

Conclusion

This thesis has presented evidence on two topics relating to organisational practices in manufacturing firms in low-income countries. First, I examined how managerial practices can foster bottom-up innovation in Chapter 1. Second, I explored determinants of differences in performance both within and across firms, and how such differences can be reduced in a cost-effective way in Chapter 2 and 3.

In Chapter 1, I find that managerial practices in the form of structured peer-to-peer lunches significantly improve the quantity and relevance of ideas, which are implementable in the firm. Managerial practices in the form of additional encouragement from trained supervisors are not beneficial, and even come at the cost of undermining workers' critical voice about supervisors. The findings suggest that the increase in the quantity and quality of submitted ideas is driven by changes in workers' aspirations and ease of submission through the lunch sessions. Further, I find evidence of a peer-mentor effect: less-experienced workers benefit from being seated opposite a more-experienced friend during lunch. Finally, I show that a perceived deterioration in the voice climate drives the undermined critical voice of workers assigned to receive supervisor encouragement. Whilst the interventions provide a first step to understand impediments to bottom-up innovation, there are a number of potential extensions to this project. Exploring the mechanisms further is a promising avenue for future research. One could think about specific interventions targeting workers' aspirations directly, e.g. through the display of winning ideas of role models. Further work could also probe more deeply into possible ways to effectively engage middle managers into the suggestion process in greater depth. Besides, I did not attempt to look at the long-run effects of a successful implementation of the ideas. Using textual analysis, one could also look closer into

the similarity of suggestions of seating partners to learn more about communication among co-workers. In addition, the experiment could be extended to cover a longer time period, which would allow for a more detailed analysis of learning effects across time as well as variations in effects after the announcement of winning ideas. For the latter, I do not find any evidence in the short-run. Finally, the theoretical framework could be extended by, for example, making it dynamic and allowing for changes in workers' efforts across time.

Chapter 2 studied the impact of a low-cost management training programme that aims to improve both firm productivity and working conditions. We analysed detailed administrative and survey data and find positive effects on worker-related outcomes, such as worker wellbeing. Practices aimed at improving communication on the production floor are found to be more effective than technical and production-related practices. While we do not see any improvements in productivity, we find decreases in exit rates, turnover and absenteeism. There are a number of avenues for further research on this topic. Future work could examine the mechanisms that explain the successful implementation of human-resource related practices compared to technical-related practices more closely. For example, future interventions could directly address sources for delays or failures of decision-making processes, such as insufficient cross-departmental communication and coordination. While we find a low degree of complementarity of generic practices, further work could explore the complementarities of practices in greater depth. It would also be interesting to study which of the implemented practices have been kept by managers in the long-run and to evaluate resulting effects. Survey results indicate that at least some of the practices were still in place six months after the initial implementation, but our experiment did not allow us to validate this finding for a longer time horizon. Finally, this research could also be expanded to other industries and settings.

While Chapter 1 and 2 provided experimental evidence from two randomised controlled trials, Chapter 3 uses non-experimental data to shed light on productivity differences within firms and across time. We find that performance differences across production units can be traced back to varying degrees of autonomy exercised by production managers when deciding to deviate from pre-determined

plans. Comparing the efficiency costs in the production data with efficiency costs embedded into the production plans, we find that the change in the factory's production model imposed by industry demand undermined the planning departments' ability to generate adequate production plans. We argue that deviating from plans to re-organise production within the units' organisational groups was fundamental in keeping up with competitors in a high-paced industry. There remain a number of unanswered questions and avenues for further research on this topic. We did not delve into other factors which could potentially explain heterogeneous performance across units, such as the unavailability of fabric or differences in the skill set of the workforce. Our datasets unfortunately lack information on the capabilities of both workers and managers, so we were also not able to link the degree of autonomy of departments to managers' capabilities. Further work, ideally with more detailed data on the distribution of skills, could explore many of these remaining questions. There is also scope to examine further internal and external determinants of production, such as external pressure imposed by the proximity of delivery dates. Moreover, further work could go into more depth on the analysis of the deviations between planned and actual production. Exploring additional ways to compare datasets from different departments and to analyse what they teach us about communication and coordination within firms using statistical methods is an exciting avenue of research, particularly as more firm-level data become available.

References

Adhvaryu, Achyuta, Teresa Molina, and Anant Nyshadham, "Expectations, Wage Hikes, and Worker Voice: Evidence from a Field Experiment," *NBER Working Paper 25866*, 2019.

Aghion, Philippe and Jean Tirole, "Formal and Real Authority in Organizations," *The Journal of Political Economy*, 1997, *105* (1), 1–29.

Almunia, Miguel, Jonas Hjort, Justine Knebelmann, and Lin Tian, "Strategic or Confused Firms? Evidence from "Missing" Transactions in Uganda," *Working Paper*, 2019.

Anderson, Michael L, "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," 2008, *103* (484), 1481–1495.

Aoki, Masahiko, "Toward an Economic Model of the Japanese Firm," *Journal of Economic Literature*, 1990, *28* (1), 1–27.

Arrow, Kenneth J, *The Limits of Organization*, Norton, 1974.

Atkin, David, Azam Chaudhry, Shamyla Chaudry, Amit Khandelwal, and Eric Verhoogen, "Organizational Barriers to Technology Adoption: Evidence from Soccer-Ball Producers in Pakistan," *Quarterly Journal of Economics*, 2017, *123* (3), 1101–1164.

Baird, John E, *The Dynamics of Organizational Communication*, Harper & Row, 1977.

Bandiera, Oriana, Andrea Prat, and Tommaso Valletti, "Active and Passive Waste in Government Spending: Evidence from a Policy Experiment," *American Economic Review*, 2009, *99* (4), 1278–1308.

_ , _ , Stephen Hansen, and Raffaella Sadun, "CEO Behavior and Firm Performance," *Journal of Political Economy*, 2020, *128* (4).

Bartelsman, Eric J and Mark Doms, "Understanding Productivity: Lessons from Longitudinal Microdata," *Journal of Economic Literature*, 2000, *38* (3), 569–594.

Battiston, Diego and Tom Kirchmaier, "Face-to-Face Communication in Organisations," *Working Paper*, 2018, pp. 1–62.

BBC, "Eid al-Adha: Why are there two Eids?," May 2019. Library Catalog: www.bbc.uk Section: Newsround.

Benjamini, Yoav, Abba M Krieger, and Daniel Yekutieli, "Adaptive linear step-up procedures that control the false discovery rate," *Biometrika*, 2006, *93* (3), 491–507.

BGMEA, Bangladesh Garment Manufacturers and Exporters Association, "Trade Information," July 2020. Library Catalog: https://www.bgmea.com.bd Section: Trade.

Bloom, Nicholas and John Van Reenen, "Measuring and Explaining Mangement Practices Across Firms and Countries," *Quarterly Journal of Economics*, 2007, *122* (4), 1351–1408.

_ , Aprajit Mahajan, David McKenzie, and John Roberts, "Do Management Interventions Last? Evidence From India," *World Bank Policy Research Working Paper no.8339*, 2018.

_ , Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts, "Does managment matter? Evidence from India," *The Quarterly Journal of Economics*, 2013, *128* (1), 1–51.

_ , John Van Reenen, and Raffaela Sadun, "The Organization of Firms across Countries," *The Quarterly Journal of Economics*, 2012, *127* (4), 1663–1705.

_ , Raffaella Sadun, and John Van Reenen, "Management as a Technology?," *NBER Working Paper 22327*, 2017.

_ , Renata Lemos, Raffaela Sadun, Daniela Scur, and John Van Reenen, "The New Empirical Economics of Management," *Centre for Economic Performance*, 2014, *41.*

Bolton, Patrick and Mathias Dewatripont, "The Firm as a Communication Network," *The Quarterly Journal of Economics*, 1994, *109* (4), 809–839.

Boudreau, Laura, "Multinational enforcement of labor law: Experimental evidence from Bangladesh's apparel sector," *Working Paper*, 2020.

Bruhn, Miriam, Dean Karlan, and Antoinette Schoar, "The Impact of Consulting Services on Small and Medium Enterprises : Evidence from a Randomized Trial in Mexico Antoinette Schoar," *Journal of Political Economy*, 2018, *126* (2).

Burris, Ethan R, Evan A Bruno, James R Detert, and Andrea Caldwell, "From Barriers to Drivers : Understanding Different Approaches to Getting Systematic Upward Voice in Organizations," *The Oxford Handbook of Organizational Citizenship Behavior*, 2018.

Bursztyn, Leonardo and Robert Jensen, "Social Image and Economic Behavior in the Field: Identifying, Understanding, and Shaping Social Pressure," *Annual Review of Economics*, 2017.

Cai, Jing and Shing-Yi Wang, "Empowering Workers: Evidence from Auto Manufacturing in China," *Project*, 2018.

_ and Shing yi Wang, "Improving Management through Worker Evaluations: Evidence from Auto Manufacturing," *NBER Working Paper 27680*, 2020, pp. 1–54.

Chatterji, Aaron and Duke Fuqua, "When Does Advice Impact Startup Performance?," *Harvard Business School Working Paper no. 17-100*, 2018.

Coffman, Katherine Baldiga, "Evidence on self-stereotyping and the contribution of ideas," *The Quarterly Journal of Economics*, 2014, pp. 1625–1660.

Cottini, Elena, Takao Kato, and Niels Westergaard-nielsen, "Adverse workplace conditions , high-involvement work practices and labor turnover: Evidence from Danish linked employer-employee data," *Labour Economics*, 2011, *18* (6), 872–880.

Dalton, Patricio S, R Julius, and Bilal Zia, "Learning Business Practices from Peers: Experimental Evidence from Small-scale Retailers in an Emerging Market," 2018.

Dellavigna, Stefano, John A List, Ulrike Malmendier, and Rao, "Estimating Social Preferences and Gift Exchange at Work," *NBER Working Paper No. 22043*, 2016.

_ , U C Berkeley, and Devin Pope, "What Motivates Effort? Evidence and Expert Forecasts," *Review of Economic Studies*, 2018, (85), 1029–1069.

Edmondson, Amy, "Psychological Safety and Learning Behavior in Work Teams," *Administrative Science Quarterly*, 1999, *44*, 350–383.

_ and Zhike Lei, "Psychological Safety: The History, Renaissance, and Future of an Interpersonal Construct," *The Annual Review of Organizational Psychology and Organizational Behavior*, 2014, *1*, 23–43.

Friebel, Guido and Michael Raith, "Abuse of Authority and Hierarchical Communication Abuse of authority and hierarchical communication," *The RAND Journal of Economics*, 2004, *35* (2), 224–244.

Gant, Jon, Casey Ichniowski, and Kathryn Shaw, "Social Capital and Organizational Change in High-Involvement and Traditional Work Organizations," 2002, *11* (2), 289–328.

Garicano, Luis and Esteban Rossi-Hansberg, "Knowledge-Based Hierarchies: Using Organizations to Understand the Economy," *Annual Review of Economics*, 2015, *7* (1), 1–30.

_ and Luis Rayo, "Why Organizations Fail: Models and Cases," *Journal of Economic Literature*, 2016, *54* (1), 137–192.

Geanakoplos, John and Paul Milgrom, "A Theory of Hierarchies Based on Limited Managerial Attention," *Journal of the Japanese and International Economies*, 1991, *5*, 205–225.

Gibbons, Robert and Rebecca Henderson, "What Do Managers Do? Exploring Persistent Performance Differences among Seemingly Similar Enterprises'," *Harvard Business School Division of Research Working Papers*, 2012, *13-020.*

Gibbs, Michael, Susanne Neckermann, and Christoph Siemroth, "A field experiment in motivating employee ideas," *The Review of Economics and Statistics*, 2017, *99* (4), 577–590.

Giorcelli, Michela, "The Long-Term Effects of Management and Technology Transfers," *American Economic Review*, 2019, *109* (1), 121–152.

Gneezy, Uri and John A List, "Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments," *Econometrica*, 2006, *74* (5), 1365–1384.

Grant, Robert M, "Toward a Knowledge-Based Theory of the Firm," *Strategic Management Journal*, 1996, *17*, 109–122.

Hahn, Youjin, Asadul Islam, Eleonora Patacchini, and Yves Zenou, "Friendship and Female Education: Evidence from a Field Experiment in Bangladeshi Primary Schools," *The Economic Journal*, 2019, *130* (627), 740–764.

Hanna, Rema, Joshua Schwartzstein, and Sendhil Mullainathan, "Learning Through Noticing: Theory and Evidence From a Field Experiment," *The Quarterly Journal of Economics*, 2014, pp. 1311–1353.

Hayek, Author F A, "The Use of Knowledge in Society," *The American Economic Review*, 1945, *35* (4), 519–530.

Hendel, Igal and Yossi Spiegel, "Small Steps for Workers, a Giant Leap for Productivity," *American Economic Journal: Applied Economics*, 2014, *6* (1), 73–90.

Higuchi, Yuki, Hoang Vu Nam, and Tetsushi Sonobe, "Management skill, entrepreneurial motivation, and enterprise survival: Evidence from randomized experiments and repeated surveys in Vietnam," *Working Paper*, 2017.

Hjort, Jonas, "Ethnic divisions and production in firms," *The Quarterly Journal of Economics*, 2014, pp. 1899–1946.

Hsieh, Chang-Tai and Peter J Klenow, "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, 2009, *124* (4), 1403–1448.

Iacovone, Leonardo, William Maloney, and David McKenzie, "Improving Management with Individual and Group-Based Consulting: Results from a Randomized Experiment in Colombia," *Working Paper*, 2018.

Jäger, Simon, Benjamin Schoefer, and Jörg Heining, "Labor in the Boardroom," *Working Paper*, 2020.

Kling, Jeffrey R, Jeffrey B Liebman, and Lawrence F Katz, "Experimental Analysis of Neighborhood Effects," *Econometrica*, 2007, *75* (1), 83–119.

Kondracki, Chris, "NextGen Supply Chain: H&M's Low Tech Approach to Inventory Management," April 2018. Library Catalog: www.mmh.com/article.

Lafortune, Jeanne, "Role models or individual consulting : The impact of personalizing micro-entrepreneurship training," *American Economic Journal: Applied Economics*, 2018, *10* (4), 222–45.

Levenshtein, Vladimir Iosifovich, "Binary Codes Capable Of Correcting Deletions, Insertions, and Reversals," *Soviet Physics Doklady*, 1966, *10* (8), 845–848.

Levitt, S. D., John List, and Chad Syverson, "Toward an Understanding of Learning by Doing: Evidence from an Automobile Assembly Plant," *Journal of Political Economy*, 2013, *121* (4), 643–681.

Lillico, T M, *Managerial Communication*, New York: Pergamon Press, 1972.

McAfee, Preston R and John McMillan, "Organizational Diseconomies of Scale," *Journal of Economics & Management Strategy*, 1995, *4* (3), 399–426.

McKenzie, David and Christopher Woodruff, "What Are We Learning from Business Training and Entrepreneurship Evaluations around the Developing World?," *World Bank Research Observer*, 2014, *29* (1), 48–82.

McKinsey & Company, "Is Apparel Manufacturing coming home? Nearshoring, Automation, and Sustainability – Establishing a Demand-focused Apparel Value Chain," October 2018. Library Catalog: www.mckinsey.com/industries Section: Retail.

Melumad, Nahum, Dilip Mookherjee, and Stefan Reichelstein, "A Theory of Responsibility Centers," *Journal of Accounting and Economics*, 1992, *15*, 445–484.

Menzel, Andreas, "Knowledge Exchange and Productivity Spill-overs in Bangladeshi Garment Factories," *Working Paper*, 2020.

Morrison, Elizabeth W, "Employee Voice and Silence," *The Annual Review of Organizational Psychology and Organizational Behavior*, 2014, *1*, 173–197.

Paton, Elizabeth, "After Factory Disaster, Bangladesh Made Big Safety Strides. Are the Bad Days Coming Back?," March 2020. Library Catalog: www.nytimes.com Section: Asia.

Prat, Andrea, Jacques Crémer, and Luis Garicano, "Language and the Theory of the Firm," *The Quarterly Journal of Economics*, 2007, *122* (1), 373–407.

Qian, Yingyi, "Incentives and Loss of Control in an Optimal Hierarchy," *The Review of Economic Studies*, 1994, *61* (3), 527–544.

Rigtering, J P Coen, G Utz Weitzel, and K Katrin Muehlfeld, "Increasing quantity without compromising quality: How managerial framing affects intrapreneurship," *Journal of Business Venturing*, 2019, *34* (2), 224–241.

Sandvik, Jason J, Richard E Saouma, Nathan T Seegert, and Christopher T Stanton, "Workplace knowledge flows," *The Quarterly Journal of Economics*, 2020, pp. 1–46.

Schwarz, Carlo, "lsemantica: A Stata Command for Text Similarity based on Latent Semantic Analysis," *The Stata Journal*, 2019, *19* (1), 129–142.

Smith, Noah, "Management Consultants Might Be the Best Foreign Aid," July 2019. Library Catalog: www.bloomberg.com Section: Opinion.

Sting, Fabian, Christoph Fuchs, Maik Schlickel, and Oliver Alexy, "How to Overcome the Bias We Have Toward Our Own Ideas," *Harvard Business Review*, 2019.

Syverson, Chad, "Market Structure and Productivity: A Concrete Example," *NBER Working Paper 10501*, 2004.

_ , "What Determines Productivity?," *Journal of Economic Literature*, 2011, *49* (2), 326–365.

The Economist Intelligence Unit, "Bangladesh economy: Challenges lie ahead for the Garment Industry," March 2019. Library Catalog: https://country.eiu.com Section: Bangladesh.

The World Bank, "World Bank National Accounts data, and OECD National Accounts data files," 2020. Library Catalog: https://data.worldbank.org/indicator Section: GDP per capita growth - Bangladesh.

Thornton, R. A. and P. Thompson, "Learning from Experience and Learning from Others: An Exploration of Learning and Spillovers in Wartime Shipbuilding," 2001.

Tonin, Mirco and Michael Vlassopoulos, "Disentangling the sources of pro-socially motivated effort: A field experiment," *Journal of Public Economics*, 2010, *94* (11-12), 1086–1092.

_ and _ , "Corporate Philanthropy and Productivity: Evidence from an Online Real Effort Experiment," *Management Science*, 2015, *61* (8), 1795–1811.

Uckat, Hannah and Christopher Woodruff, "Learning How to Choose or Learning How to Lead? Experiments on Selecting and Training Female Managers in Bangladesh's Garment Industry," *Working Paper*, 2020.

Walker, Francis A, "The Source of Business Profits," *Quarterly Journal of Economics*, 1887, *1* (3), 265–288.

Weber, Roberto A and Colin F Camerer, "Cultural Conflict and Merger Failure: An Experimental Approach," *Management Science*, 2003, *49* (4), 400–415.

Williams, Martin J and Liah Yecalo-Tecle, "Innovation, Voice, and Hierarchy in the Public Sector: Evidence from Ghana's Civil Service," *Economic Development & Institutions Working Paper Series 19/09*, 2019.

WVS, "World Values Survey Wave 7 (2017-2020)," 2020. Library Catalog: www.worldvaluessurvey.org Section: Data & Documentation, WVS Wave 7.