

DISCUSSION PAPER SERIES

IZA DP No. 14146

**World Management Survey at 18:
Lessons and the Way Forward**

Daniela Scur
Raffaella Sadun
John Van Reenen
Renata Lemos
Nicholas Bloom

FEBRUARY 2021

DISCUSSION PAPER SERIES

IZA DP No. 14146

World Management Survey at 18: Lessons and the Way Forward

Daniela Scur
Cornell University

Raffaella Sadun
Harvard Business School

John Van Reenen
LSE, MIT and IZA

Renata Lemos
World Bank

Nicholas Bloom
Stanford University and IZA

FEBRUARY 2021

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

World Management Survey at 18: Lessons and the Way Forward*

Understanding how differences in management “best practices” affect organizational outcomes has been a focus of both theoretical and empirical work in the fields of management, sociology, economics and public policy. The World Management Survey (WMS) project was born almost two decades ago with the main goal of developing a new systematic measure of management practices being used in organisations. The WMS has contributed to a body of knowledge around how managerial structures, not just managerial talent, relates to organizational performance. Over 18 years of research, a set of consistent patterns have emerged and spurred new questions. We will present a brief overview of what we have learned in terms of measuring and understanding management practices and condense the implications of these findings for policy. We end with an outline of what we see as the path forward for both research and policy implications of this research programme.

JEL Classification: L2, M2, O14, O32, O33

Keywords: management practices, policy toolkit, world management survey, productivity

Corresponding author:

John Van Reenen

MIT

77 Massachusetts Ave

Cambridge, MA 02139

USA

E-mail: vanreene@mit.edu

* Daniela Scur is the lead author, and subsequent order of names were randomized. We thank Margaret Stevens, an anonymous referee and the participants of the authors’ workshop in Oxford for their helpful feedback, and Fabiano Dal Ri for excellent research assistance.

1 Introduction

Social scientists have studied the importance of managers and of management practices for over a century, and today armed with large and representative datasets are confident that management is crucial for organizational performance. Understanding how differences in the adoption of management “best practices” affect organizational outcomes has been a focus of both theoretical and empirical work in the fields of management, sociology, economics and public policy. Earlier efforts relied primarily on case studies, but over the past two decades we have seen an explosion of new datasets and computing power that allow for careful measurement of these practices across a range of organizational types.

The World Management Survey (WMS) project was born almost two decades ago with the main goal of developing a new systematic measure of management practices being used in organisations. Since 2002, we have been building the largest cross-country dataset including manufacturing and retail firms, schools and hospitals, and we have provided assistance to other teams looking to use the methodology in numerous other industries. To date, our dataset includes over 13,000 firms and 4,000 schools and hospitals, spanning over 35 countries.¹ Other teams using the WMS methodology have collected data on government bureaucracies [Rasul and Rogger, 2018, Rasul et al., 2018], universities [McCormack et al., 2014], primary healthcare facilities and many more. The methods and availability of WMS data is open to all, and at time of writing over 5,000 researchers have downloaded the public dataset for research projects and benchmarking exercises.

The WMS has contributed to a body of knowledge around how managerial structures, not just managerial talent, relates to organizational performance. A series of stylized facts have taken form over the past eighteen years of research, both elucidating the robust positive performance-management relationship, understanding how to improve measurement (such as the more recent Management and Organizational Practices Surveys) and raising a number of important questions about mechanisms driving the diversity of management practices. We present a brief overview of what we have learned in terms of measuring and understanding management practices and condense the implications of these findings for policy in the form of a simple table: a “management policy toolkit”. We end with an outline of what we see as the path forward for both research and policy implications of this research programme.

2 What have we learned thus far

We start by taking stock of what we have learned from the WMS about measurement and the key relationships between management and performance.

¹At time of writing, the latest publicly available dataset includes 35 countries.

2.1 Measurement

2.1.1 Original WMS

Prior to the development of the WMS, research on management was often focused on small (and often, non-representative) numbers of firms in single countries. Case studies remain extremely popular, and while much can be learned from these “clinical studies” for formulating theories and considering mechanisms, it is hard to generalize from them. The typical case study has a sample size of one and is inherently selected. The goal of the WMS project is to systematically collect data on the types of practices used at thousands of organisations, across industries, under different settings and over time, while maintaining comparability.

The original WMS was set up as an interview-based survey tool, where highly trained interviewers engaged a middle manager in a semi-structured conversation about the day-to-day practices followed at their establishment [Bloom and Van Reenen, 2007]. The respondent managers were those who were senior enough in their establishment to have decision powers, but not too senior so as to be detached from the day-to-day running of the establishment. For example, manufacturing plant managers, retail shop managers, hospital department nurse managers and school principals or head teachers. It is set up as an interview, and the questions although structured, are mostly open-ended so the manager being interviewed is not guided towards what a high or low scoring answer might be. The advantage of this is that it avoids the manager simply giving the answer she thinks the interviewer wants to hear. The conversation follows a set of broad practices spanning operations/monitoring, target setting and people/incentive management practices. Each broad sector has between 18 and 20 topics and each is scored on a scale of 1 (little to no structure/”weak practices”) to 5 (well-structured/”best practices”).

The main benefit of this methodology is the many steps ingrained in the data collection process that yield high quality data. As independent interviewers assign the scores, it is feasible to build a comparable dataset across countries and industries that mitigates biases borne out of self-response. For example, common issues such as the mood of the manager, cultural expectations dictating “right or wrong” answers, different understanding of the question text and survey fatigue are not as salient in this method. The method is double blind on the side of the interviewees, but also the interviewers who will not typically know anything about the organization in advance. The average WMS response rate is usually between 40 and 50 percent, which is extremely high considering that many firms surveys typically get far lower response rates.²

The main drawback, however, is the cost. Running a full survey usually costs an average of \$300 to \$500 per interview. This is because running a full wave of the original survey entails large fixed costs and even larger personnel costs. The fixed costs include a large room equipped with desks, chairs, computers, phones, headsets as well as the survey software, calling plans and

²For example, Altig et al. [2020], Ben-David et al. [2013], Bloom et al. [2019b], Bartik et al. [2020] where response rates in firms survey range from 0.1% to 13%.

other phone charges. The personnel costs are also substantial. As the interviews are run on the phone, it requires recruitment and training of interviewers who can understand and internalize the methodology well enough to ask open questions, quickly synthesize the manager’s answers to ask follow up questions and make a judgement of the score the answers warrant. They also need to have good phone manners and charisma to keep managers engaged in the interview for at least 45 minutes, with only a promise of a report in return. As such, these are not easy workers to recruit and warrant high wages. For each 5 interviewers, there is also the additional cost of a calibration supervisor who listens to a large share of calls and scores alongside the interviewer for later comparison (this process is called “double scoring”).³

The high cost of this method is the main constraint to expanding the sample size. With an eye to scalability, other methods have arisen that are closer to traditional survey approaches of fixed scripts and closed answers and require minimal training or supervision. We briefly describe them in turn, and summarize the pros and cons of the various methods in Figure 1.

2.1.2 Semi-closed, closed and self-respondent methods

The Management and Organizational Practices Survey (MOPS) was born out of a partnership with the US Census Bureau, where a new module for the Annual Survey of Manufacturers in reference year 2010 included a set of questions that roughly mirrored those that the WMS measured. The module was sent to over 40,000 establishments across the United States and achieved an 85% response rate [Bloom et al., 2019a].⁴ Since then, similar exercises have been replicated in Australia, China, Colombia, Denmark, Finland, Germany, Japan, Mexico, the Netherlands, Pakistan and the UK. The main benefit of this method is the scalability and potentially lower cost of operations, though a major drawback is the difficult of comparing these management scores across countries (e.g. we can easily compare management scores from the Office of National Statistics across UK firms, but less easily to management scores from the Census on US firms). When done in partnership with government data agencies (such as the US Census, Mexico’s INAGI, or Pakistan’s State Bank and Bureau of Statistics), however, it can be powerful for large national sample sizes and has a major benefit of additional linkages with other administrative datasets.

Two intermediate approaches sit between the self-respondent census-style method and the original phone WMS with open-ended questions. One is to use phone or hand delivery of closed-ended questionnaires with the goal of mitigating low response rates. This was the approach in Bloom et al. [2012c] to survey Eastern European and Eurasian firms via the phone, the second wave of the Pakistan MOPS included hand delivery of the questionnaires. The benefit is that the level of training and monitoring required of interviewers is lower, enabling surveys to be conducted in difficult to reach areas in person or by phone.

³The WMS trainers have also trained multiple waves across years, and help back-score a random set of surveys to further help with calibration.

⁴Answering this survey is mandatory in the United States.

Another intermediate approach is to keep the same set of open-ended questions but use a much more detailed scoring grid that almost reaches a checkbox-based scoring rubric. This method is particularly useful in developing countries, as it allows for finer measurement of the often-thicker lower tail of badly managed establishments. Additionally, this method does not necessarily require leveraging highly trained interviewers to collect data: the scoring rubric is detailed enough to remove room for interpretation. As such, the method is dubbed the “Development WMS”. It was piloted with primary schools in Andhra Pradesh (see [Lemos et al. \[2021\]](#) for full details), and subsequently applied in Puerto Rico, Colombia, Mexico, Haiti, Pakistan and Tanzania.⁵

2.1.3 Leveraging existing non-WMS questionnaires

A more recent approach looks to leverage existing large datasets that include questions on the management topics being measured by the WMS. For example, in education, standardized tests are often administered to all schools within a country, or at least to a representative sample with appropriate weights available to researchers as part of the database. These standardized tests usually also feature a principal questionnaire (and sometimes student and teacher questionnaires as well) including questions that reflect some of the topics surveyed in the WMS. [Leaver et al. \[2019\]](#) developed the method using data from the OECD’s Programme for International Student Assessment (PISA) as well as the Brazilian Prova Brasil and validated the scores against the WMS existing dataset. These Education Survey Management Indices (ESMI) allow for a comparison across a much larger number of countries and schools, though the downside is that questionnaires sometimes change over time rendering comparisons more challenging.

2.1.4 Caveats

While different methods solve different shortcomings of the original methodology, they all still share some caveats that are important to highlight.

What is “good management”? The focus of the WMS questions (and indeed the topics in these alternative indices) are based on a set of practices that were ex-ante likely to be causal determinants of better performance. The design was based on a consensus of many consultants and industry experts on what core factors were important in the efficiency of a good or service that a firm was delivering. While this has been verified to be true empirically, the measure does not purport to encompass the entire spectrum of “what is management”. It does include fairly non-controversial topics where there is a consensus in terms of “good” and “bad” practices, from having a reasonable number of key performance indicators and keeping track of them regularly to having a system in place to identify which workers are performing well and which ones need help to improve.

⁵The set of questions used in Pakistan, Tanzania and Puerto Rico were a smaller subset from the original 20.

Our WMS management evaluation, however, leaves out aspects such as innovation, finance, pricing, marketing, and opening and closing decisions (among many others). In the public sector, it leaves out topics like funding strategies, management of school safety and of student socio-emotional development. While these remain crucial for the success of these organizations, they are less straightforward to judge on a strictly increasing scale as the rest of the management practices included in the indices.⁶ Further, the monotonic scale means we may miss some crucial variation in informal practices that managers may carry out in formally constrained environments, such as heavily unionized firms or public organizations like schools or hospitals.

At its core is the question of whether these “basic” practices are complements or substitutes with other more “strategic” aspects. It may well be that a particular type of firm, say, one focused on the design of highly innovative products, would do worse with strict targets and metrics at every step of production. On the other hand, efficient data collection and value mapping coupled with appropriate incentives could make an R&D lab more efficient at churning innovative products.⁷ From the worker side, on the one hand, a focus on efficiency and formality in targets could make for a stressful working environment, but on the other hand perhaps formal indicators and transparent performance evaluations attract good workers and minimize discriminatory practices.⁸

These are all important empirical questions. While not part of the core measurement, some of these aspects are measurable in other ways and we review the evidence of the relationship between basic management practices and various organizational and productivity outcomes in the next section.

Is “good management” universal? Another concern with the focus of the questions covered in these indices is that they are culturally biased towards “Anglo Saxon” values or types of practice that would fail to reflect a real existing relationship with performance. Over the years of data collection across a large number of culturally diverse countries, we have found that this is not a major issue. First, we implement methodological steps to ensure interviewers have exposure to firms from multiple countries by interviewing firms in their target country as well as other countries for which they have the language skills.⁹ Second, the practices measured are fairly core and “simple” (that is, not strategic), such that they are arguably less susceptible to cultural differences. Third, we can empirically test whether the relationship between these practices and performance is different across countries (it is not), and whether multinational firms belonging to different home countries have different practices when operating in foreign countries (they generally do not).

⁶To be sure, there are now new initiatives looking to explore these less straightforward aspects, including a World Strategy Survey and an Educational Financial Management Survey.

⁷For example, IDEO is a design firm that is known for its well-defined routines for design thinking [Brown, 2008].

⁸Indeed, the limited evidence we have on this suggests firms with better management practices have better work-life balance, including higher-wages, lower employee turnover and better subjective work-life balance survey responses (e.g. Bloom et al. [2009]).

⁹For example, many interviewers would also interview firms in the UK or the US alongside interviewing firms in Ethiopia or Japan.

To be sure, the best test of whether these measures help explain useful and interesting variation across establishment is matching the data with external measures that we care about (productivity, profitability, innovation, worker welfare, wage inequality, etc.) and verifying the direction of the relationship. We review the main stylized facts we and other researchers have found in the next section.

2.2 Main stylized facts

2.2.1 Management varies across and within countries and sectors

The original WMS method yields individual measures ranging from 1 to 5 across 18 different topics in manufacturing, 19 in retail, 21 in hospitals and 23 in schools. We briefly describe the topics and provide a set of example questions in Tables 3 to 6. As the questions are naturally correlated, they are usually aggregated into the “overall” management index (including all questions) or into sub-indices of the questions covering each of the portions of the questionnaire (lean operations, monitoring, target-setting and people management).¹⁰ The standard way of combining the management questions used in most papers is simply standardizing each question, taking the average of the set of relevant questions, and then standardizing the average. The index value varies slightly depending on how it is constructed, but as we will review in the next section, the aggregation method does not seem to matter much for the strength of the management-performance correlation.

Exploring the questions and their relationship with each other can be interesting in itself. A factor analysis of the 18 manufacturing questions yields one principal component that explains most of the variance and also loads positively on all questions, suggesting that if a firm scores highly on one practice, it is likely to score highly on all other practices, too (Table 1). This can be interpreted as a latent variable of “good management”. The second factor explains much less of the variation, but it has interesting loadings: opposite signs for general operations questions and people management questions. We repeat this exercise for public hospitals and public schools, and the patterns are strikingly similar. Both samples also display a strong first factor that explains a large share of the total variance, and a second factor with opposite sign loadings approximately divided along general operations and people management lines. The one exception is public hospitals, where a third factor also emerges that loads positively on lean operations and people management, and negatively on monitoring and target setting. The general pattern suggests that across all sectors there are some clear lines of specialization in general operations versus people management. This could be reflecting the strength of labour regulations across countries, as some tend to be more restrictive (such as France) relative to others (such as the US).

Looking across countries, our WMS manufacturing data shows substantial differences in the average management practices index across countries as well as within countries. Across countries, management scores seem to track closely with levels of development (Figure 2). The United States

¹⁰For some purposes, aggregating all the non-people management questions into “general operations” is appropriate. For example, in Bloom et al. [2015b] or in Cornwell et al. [2020].

has the highest average management score followed by other high-income countries such as Japan, Sweden and Canada. Southern European countries such as Spain and Greece have lower average scores, similar to upper-middle-income countries such as Chile and Mexico. Emerging economies such as India and China are next, with the African countries in our sample coming last on the ranking. As the survey eligibility sets the minimum firm size at 50 employees, this is not driven by low-income countries having smaller firms. In fact, the size of the circles corresponds to the relative median firm size within the country and they are similar to each other across the sample of countries.¹¹ But this simple ranking masks substantial within-country heterogeneity.

Figure 3 plots firm level kernel density curves for the 35 countries of the two sub-indices of management, operations (including core operations, monitoring and target-setting) and people management. Two distinguishing features of the score leader, the United States, are that it has an unusually thin left tail of low scoring firms, and the distributions of operations and people management are not too different from each other. It is worth highlighting that even in the countries that are closer to the bottom of that average ranking, there are firms that are able to adopt these practices and do well, so while it is unlikely that something structurally blocks these practices from being implemented, there certainly are barriers to their diffusion.

We find a mostly similar pattern in the public sector (schools and hospitals), though the United States is not the leader in the management of public schools. Figure 4a shows the ranking for hospitals and Figure 4b shows the ranking for schools. While each industry has a set of industry-specific “core operations” questions, there is a large set of overlapping questions covering monitoring, target-setting and people management practices. Figure 5a shows the distribution of scores within each of the countries for which we have all three sectors. On average, scores for hospitals are lower relative to manufacturing, and schools are lower than hospitals in all countries in our sample. The deficit is similar across the operations management and people management questions, too. In all, variation across countries is interesting to observe and important to document. But does this difference in management practices matter?

2.2.2 Management matters

There is a large literature documenting the importance of management practices in organizations, much of which will be reviewed in the papers in this issue. As such, we will focus on the lessons we learned from the WMS (and WMS-like) measures. The various management indices have been used to explore the relationship between management and establishment performance (productivity, student achievement, mortality rates), as well as other organizational outcomes (work-life balance, wage inequality).

Firms The vast majority of studies using the WMS measures are not randomized control trials (RCTs), but rather use non-experimental regressions with cross-sectional or panel data. Figure

¹¹We have substantially fewer countries in the retail dataset, but the pattern is consistent.

6 shows the average value of each of six performance outcomes across the deciles of the main management measure. The relationship is monotonically positive: firms in the higher deciles of management have higher productivity, operating profit, output growth, exports, Research and Development expenditures and patents [Bloom et al., 2019a]. This positive relationship has been consistently verified across a number of studies and settings. Figure 7 summarizes the coefficients and confidence intervals from these studies.

It is reassuring that the core correlation is robust to so many different (and demanding) specifications and contexts. The range of specifications include firms in all available countries and running OLS regressions in levels with the cross-sectional data as well as fixed effects regressions with panel data. Much of the pooled country analysis comes from semi-public performance data from public reports data aggregators such as Bureau van Dijk. This could mean a selection bias towards countries with strong reporting requirements (as in Europe), but a number of papers also match the WMS data to country-specific confidential administrative datasets and continue to find the same patterns.¹² The core management-performance results for the WMS data with matching public performance data is robust to this aggregation, or alternative ones using factor analysis or the more recent Anderson [2008] index.¹³

Still, non-experimental evidence has limits, including the omnipresent omitted variable bias and reverse causality issues. Even in panel data analysis including time-invariant fixed effects, estimates could be susceptible to time-variant unobservables that are correlated with both management and performance. In cases like this, RCT evidence can be a powerful complement to broad-based correlational evidence and Bloom et al. [2013] is an example of this. They started with a set of textile manufacturers in India (a simple technology that is easy to compare across firms), and divided them into two groups: one that was provided with high quality management consulting to help them adopt the type of modern management practices measured by the WMS (the treatment group), and another that was not provided with consulting help but only data collection visits (the control group). The experiment showed a causal effect of adopting these practices on productivity that mirrored the magnitudes seen in the correlational work: one standard deviation increase in management quality led to approximately 10% increase in productivity. A subsequent follow-up several years after the experiment found that the impacts were lasting: while about half of the practices originally adopted were still in use, they had spread across untreated plants within the same firm — though notably not across firms [Bloom et al., 2020b].

But productivity is not the only thing management practices affect. Researchers have also found a positive relationship between management practices and family friendly policies and work-life balance [Bloom et al., 2009], energy efficiency [Bloom et al., 2010], good hiring and firing decisions as well as within-firm wage inequality [Cornwell et al., 2020, Bender et al., 2018, Song

¹²For example, Lemos and Scur [2019] and Cornwell et al. [2020] for Brazil, Bloom et al. [2019a] for the US, and Heyman et al. [2019] for Sweden.

¹³Meagher [2013] use Bayesian techniques with WMS data from the UK, US, France and find that there is some convexity in this relationship for high scores.

et al., 2018] and tax planning [Bilicka and Scur, 2020].

Public sector The positive relationship between management and performance is also present in the public sector. In hospitals, better management practices were correlated with better patient outcomes (survival rates and lower length of stay) [Bloom et al., 2015c, 2020a]. In for-profit nursing homes, better managed is associated with lower likelihood of bad performance (although this does not hold in not-for-profit nursing homes. In schools, there is again a positive relationship between management practices and student test scores [Bloom et al., 2015b, Lemos et al., 2021], in both public and private schools. In university departments, McCormack et al. [2014] finds better management is correlated with better teaching scores and research outcomes. In government services (the post office), better management is correlated with higher efficiency at the county level [Chong et al., 2014]. In bureaucracies, Rasul and Rogger [2018] find that while the autonomy of bureaucrats is positively linked to completion rates of engineering projects, incentives and monitoring are not. As our focus is primarily on the lessons for the private sector, we abstain from a longer discussion on the public sector here and direct readers to Valero [2021] and Ali et al. [2021].

3 Potential policy directions

More generally, if we accept that there is a causal relationship between productivity and better management, there is also a macro effect of improving management and that matters for governments and government policy. Bloom et al. [2016] estimate that management accounts for, on average, about a third of the gap in total factor productivity (TFP) between the United States and other countries in the sample.¹⁴ Further, we have growing evidence that managers and management practices matter in the delivery of public sector services as well [Finizia, 2020, Bloom et al., 2015b, Rasul et al., 2018]. So what we have learned about what this means for government policy on management practices?

There is much evidence about the “drivers” of management practices, so we pull this research together in policy relevant ways. We imagine that a policy maker has a budget for management-related policies and wants advice on how the funds could be best used. We put together a “policy toolkit” summarized in Table 2. These are our subjective judgements of the literature on a core set of common policies, which we justify in the sections below.

In column 1, we split policies into “structural” which are not usually specifically targeted on management, but where research has produced evidence suggesting an indirect effect on management (generally, by reducing market frictions) and “direct” which are policies more clearly targeted on management. The rows for structural policies are competition, trade/FDI, education, labor

¹⁴For some countries, like Portugal and Italy, management can account for as much as half the difference in TFP, while for countries like Japan and Sweden it accounts for much less, closer to 10% of the gap.

regulation and governance. For direct policies, they are training (which we divide into consulting vs. formal classroom) or informational interventions. These are within-organization incremental changes rather than full system-wide reforms, focused on improving organizational or manager-specific skills and practices.

Column 2 of Table 2 includes our judgement of the strength of the evidence available for the effectiveness of each type of policy. Our decision of awarding low, medium or high is based on a combination of quantity, quality and credibility of research papers and research designs.¹⁵ Broadly, a ranking of *low* (*L*) refers to a body of evidence that primarily constitutes of cross-sectional correlational evidence. A ranking of *medium* (*M*) indicates that the evidence includes research designs with panel data, showing differences across time and some accounting of organizational fixed effects. A ranking of *high* (*H*) indicates that the evidence is arguably causal, with research designs including natural experiments or randomized control trials. We discuss a set of specific papers below, but our ranking takes into account consistency and conclusiveness of results in the literature, as well as whether it extends across contexts in determining the ultimate “strength” of the evidence.

Column 3 of Table 2 scores the net benefit of the policy, taking into account the overall benefits *for firms* minus the costs incurred by the government. In the spirit of focusing on the implications for management improvements, we abstract from making broader welfare statements that would need to take into account possible spillover effects from management. In addition, we focus only on the impact of management and the other potential benefits from these policies (e.g. general education can improve productivity and growth).

Column 4 of Table 2 considers the difficulty of implementing these policies. While some might have high net benefits, they might be “costly” in terms of political capital and thus quite difficult to implement. Our implementation difficulty assessment is based on a combination of the likelihood of push-back from interest groups or policy recipients, the financial cost of implementing a policy, and the coordination requirement to get a policy enacted (including cross-departments or across countries). A ranking of *low* (*L*) suggests we expect the policy to be a combination of politically uncontroversial, relatively cheap, and easy to coordinate. *Medium* (*M*) suggests we expect there would be some challenges to implementing the policy, though they would be tractable. For example, the policy would require some coordination, allocation of sizeable funds and/or there is some likelihood of pushback. A ranking of *hard* (*H*) suggests we expect there to be severe challenges to implementation, either from expected push-back from powerful groups, prohibitive costs of implementation or extreme coordination requirements.

Column 5 considers the time frame of when the expected effects would start materializing *after the policy has been enacted and fully implemented*. A short time frame refers to policies that would yield results within one year, a medium time frame refers to results between one and five years,

¹⁵To be sure, we are not assessing the quality of the *research papers* on the topic, but rather how useful the evidence is for the question at hand. Excellent papers sometimes can provide only limited or weak evidence on a particular topic.

and a long time frame refers to longer than five years.

3.1 Structural policy types

Product Market Competition Competition has long been recognized as an important driver of productivity, where organizations either improve their offerings to survive or exit the markets they operate in [e.g. Syverson, 2004, Galdón-Sánchez and Schmitz, 2002]. Policies include removing regulatory barriers to entry for new firms and protections for badly run incumbents. One mechanism through which competition affects productivity is management [van Reenen, 2011, Syverson, 2011]. In particular, Bloom and Van Reenen [2007] used three indicators of competition to show a strong positive relationship between higher competition and higher management scores: (i) trade openness,¹⁶ (ii) a Lerner index, where high values suggest tough competition¹⁷, (iii) the number of perceived direct competitors reported in the WMS interview. The relationship has also been found in the healthcare sector, where Bloom et al. [2015c] find a positive causal effect of competition on hospital management in the UK. In our data, we consistently find that there is a positive and significant correlation between number of perceived competitors and management scores across manufacturing, hospitals and schools.¹⁸ Overall, there is substantial strength in the evidence pool for this policy lever.

There is a very high net benefit of policies that spur competition, five out of five widgets in our ranking, as the cost is low and the benefit very high. We see it as a policy that has a medium difficulty of implementation as there may be some special interest groups that would push against dismantling protectionist barriers, but these actors are often in relatively weaker coalitions of unproductive firms [e.g. Parente and Prescott, 1999]. The time frame is “medium”, as we would expect the effect would take at least one year to start materializing, but no more than five years.

Trade and Foreign Direct Investment Trade and openness is a related policy, in that exposure to more free trade and open markets yields heightened competition. In addition, there may be benefits in terms of wider export markets, better inputs and the transplantation of better management through the affiliates of multinationals. Bloom et al. [2015a] use the timing of China’s accession to the World Trade Organization to estimate the (positive) causal impact of trade-induced competition on management both within firms and through reallocating employment towards more productive and better managed firms. Bloom et al. [2020c] show that firms with higher management scores are more likely to export, have more export destinations, export higher volumes and products of better quality. Atkin et al. [2017] randomize access to foreign markets for rug producers and find a causal evidence of learning-by-exporting, with firms showing improvements in technical

¹⁶The measure of trade openness is the ratio of imports to production at the industry-country level.

¹⁷The Lerner Index is calculated as the 1 minus the profits/sales of all other firms in the same industry and country, and high values are suggestive of low long-run profits, which happens in very competitive markets.

¹⁸For schools, however, the correlation between management and number of competitors loses significance when we control for observable characteristics though it remains positive.

efficiency. Verhoogen [2008] finds that higher export share is tied to ISO9000 certification — a set of international standards on quality measurement that overlap with a set of questions in the WMS questionnaire. At the frontier, multinational enterprises (MNE) face competition across the world and tend to have higher management scores everywhere they operate [Bloom et al., 2012b]. Further, Heyman et al. [2019] finds that differences in MNE management are an important determinant of productivity among foreign affiliates. Bloom et al. [2019a] show that multinational entry into US counties raises the management quality of local incumbent plants using “runner up” counties as a control group. They argue that these management spillovers appear to operate through the movement of managers between the entering multinational and local firms. Overall, we find that this policy lever also has a large body of high-quality evidence, so we rank the strength of evidence as high.

In terms of policy net benefit, we rank this as four and a half widgets out of five, as the benefit is almost as high as competition policy though the cost is likely to be slightly higher. Liberalizing trade may not carry direct costs, though some FDI incentives such as giving subsidies for foreign firms to set up factories and create jobs, could incur at least some financial cost. It is, however, a potentially difficult set of policies to implement as it depends on the strength of lobby and interest groups, and the set of actors that need to agree for policy to move forward. If enacted, however, we expect the time frame of these policies to be between one to five years.

General education There is a large literature on how improving the general education levels have myriad economic benefits. Using the WMS dataset and geocoded firm and university data, Feng and Valero [2020] find that firms further away from universities employ fewer skilled workers, and note that management practices are complementary with skills. They suggest the variation in the price of skills likely drives this relationships.¹⁹ Similarly, Bloom et al. [2020a] find that hospitals closer to universities offering both medical and business courses have better clinical outcomes and management practices than those further away. Queiró [2018] finds that entrepreneurs with higher levels of education start firms that are larger at entry and also grow faster. McKenzie and Woodruff [2017] survey micro and small firms across a set of emerging economies and find that those firms with owners with higher human capital tended to have better business practices. Further, the correlation between formal education of shop-floor workers and managers and management score is still present when the “worker fixed effect” (the characteristics and skills that are specific to the worker and carried with them from job to job) is accounted for, and is not country-specific: the relationship was found in both in an European rich economy like Germany [Bender et al., 2018] and an emerging economy like Brazil [Cornwell et al., 2020]. In all, we rank the strength of the evidence for this policy type as medium, because the evidence of a positive correlation is consistent across contexts and fairly good quality, but it lacks clear causal identification.

The net benefit of this policy is only two widgets out of five because education policy is costly

¹⁹Also see Valero [2021].

and the benefits *for management practices* are not as high as the cost. Implementation is not often politically controversial, but again as a result of the high cost of improving general education we rank it as medium difficulty. The time frame is undoubtedly long, as benefits would need to work their way through school cycles that could take at least five years or longer.

Labour deregulation Rigid labour laws can create barriers to adoption of management best practices, especially as they relate to people management. Bloom et al. [2019a] provide some of the best evidence to date, using the Management and Organizational Practices Survey (MOPS) from the US Census for over 35,000 firms. They show that business environment, as measured by US right-to-work (RTW) laws which generally weaken labour union power, yield higher use of incentive management practices. There is also cross-sectional evidence suggesting that labour market regulation across countries has a negative correlation with people management practices [Bloom et al., 2012a]. Further, the education sector provides additional correlational evidence as it often has public and private schools operating under more rigid and flexible labour regimes (respectively). Lemos et al. [2021] show that the difference in personnel management scores between public and private schools are larger than the difference in operations management scores, and drastically larger in emerging economies and developing countries. Overall, we rank the strength of evidence as a medium because there is at least one paper providing good causal evidence supported by others with correlational evidence, but the strong evidence comes from one particular context.

The net benefit is three widgets out of five because it is relatively cheap to adopt, but benefits seem to be concentrated on a subset of the practices that we are evaluating. It is also difficult to implement given the complexities associated with labor-market reforms. The time frame of the policy, if implemented, is likely to be medium term.

Governance Another relevant type of regulation relates to those that affect firm governance, as ownership structures and who controls the firm seems to matter for the type of practices adopted [Bloom and Van Reenen, 2007, Bloom et al., 2015d]. For example, a major policy of this type involves generous estate tax exemptions for inherited firms. Such policies can likely be credited with a larger share of second-plus generation family businesses in places such as the UK and Italy. This matters, as family firms have significantly lower management scores in every country surveyed. In fact, Lemos and Scur [2019] find that a CEO succession to a family member instead of a professional manager leads to almost a full standard deviation lower management scores. Bandiera et al. [2015] show that family firms offer flatter compensation schemes and have lower scores on management practices related to incentives. Similar to labour deregulation, we rank the strength of the evidence on this type of policy as a medium because there are a number of papers investigating various governance structures in correlational studies, but at most one paper with good causal evidence from multiple countries.

The net benefit of policies that end privileged tax treatment earns four widgets out of five, as

the policy is cheap — it could even raise government revenue through higher tax intake — and there is a high potential benefit to improving practices. However, the management benefits only materialize if the incentives succeed in pushing firms to professionalize the ranks of firm governance. It ranks as a medium to high difficulty of implementation because it will heavily depend on the political strength of some lobby groups, such as family business owners. If implemented, the time frame we would expect for policy results would be long term.

3.2 Direct Management policies

We know much less about the micro-level drivers of management practices and it is a fruitful and active research area. We see the main policy options in this space as falling in three groups: first, providing managers with training via consulting services. We define these as programs where trained consultants work with managers in the establishment to identify their main issues and how to solve them. The second group includes formal management training, via more traditional classroom-based training that the managers then take back to their firms and apply to their issues by themselves. The third group includes mixed type of information and benchmarking provision, where managers are exposed to information that allows them to compare their practices and performance to that of their peers.

Training: in-firm consulting There is substantial good evidence that in-firm consulting is highly effective for improving management practices, and we rank the strength of the evidence as high. [Bloom et al. \[2013\]](#) ran the first randomized control trial of this kind and found that a treatment of intensive, high quality management consulting yielded large improvements to practices, many of which were kept even in the long term [[Bloom et al., 2020b](#)]. This first experiment was run with textile firms in India, but the result has been consistent across other contexts including small and medium firms across a number of different countries and firm types [e.g. [Bruhn et al., 2018](#), [Karlan et al., 2015](#), [Higuchi et al., 2015](#), [Giorcelli, 2019](#), [Nam et al., 2020](#)].

The policy net benefit is three widgets out of five, as the evidence suggests the benefits are very high but the cost can also be high depending on the type of consultants used in the interventions. We rate the difficulty of implementation as medium to low because such policies are generally seen as positive and should encounter low resistance, though they do require some financial commitment. The timeline is certainly short, as we have seen the effects show up fairly consistently within one year.

Training: formal classroom A related policy tool is providing classroom-based training programs for managers, though the strength of the evidence for this policy is not as strong as for in-firm consulting. [McKenzie \[2021\]](#) discusses frontier in management training of small firms in the private sector in this issue, so will keep our comments brief. Of note are papers that specifically tackle the different modes of this type of training delivery, such as [Iacovone et al. \[2019\]](#) and [Higuchi et al.](#)

[2019]. The primary issue is the large variance in the quality and intensity of the training and the trainers, which in turn yields mixed results in terms of adoption of practices and the effects on productivity. Many papers also tend to measure outcomes in the short run, while it has since been shown that some effects of this type of training shows up in the medium run.

Unlike in-firm consulting, this type of formal training intervention has made its way into the public sector. For example, experiments such as [Fryer \[2017\]](#), where an intense management training program for principals in US public schools yield a marked improvement in student achievement. In healthcare, [Dunsch et al. \[2017\]](#) run a management training experiment with primary healthcare centres in Nigeria and find short run effects in adoption of practices as well as intermediate outcomes, but fail to see long term impacts.

The net benefit of policies that provide formal classroom training earns two widgets out of five in our ranking. This is primarily because the cost is uncertain and varies heavily with the intensity of the program and the quality of the trainers, and the benefits are also uncertain. The difficulty of implementation is likely to be low, and the time frame is medium-term.

Information and benchmarking Finally, there are a number of interventions that are adjacent to consulting and strictly formal classroom training. For example, providing information on the practices of other firms or leveraging the knowledge of local business people as mentors [[Brooks et al., 2018](#), [Cai and Szeidl, 2017](#)]. These seem to be effective in some contexts, but the evidence is limited. Providing benchmarking information also seems to be effective [[Gosnell et al., 2020](#), [Cai and Wang, 2020](#)], and indeed [Bloom et al. \[2013\]](#) report that “lack of knowledge” was an often cited reason for failing to initiate improvements to their management practices.

In the WMS questionnaire, the very last question asks the manager to score their establishment on the quality of their management practices on a scale of 1 (worst) to 10 (best). The question is set up to allow for the most honest assessment from the manager: the interviewer is clear that the reported score is meant to evaluate the average quality of the structures and practices and exclude the managers themselves, and the question comes at the end of at least one hour of discussing the practices they are meant to finally evaluate. The majority of managers over-estimate the level of the practices in their establishments, and this is particularly salient in the public sector establishments.

In all, the evidence on this type of policy initiative is still in its early stages. Our assessment of the strength of the evidence is low, not because of the quality of the existing papers but rather because of the limited scope in terms of context and small quantity of the existing research. For example, evidence from a large-scale program in India providing a mix of information and targeted support to improving school performance plans shows no impact on management practices nor student achievement [[Muralidharan and Singh, 2020](#)]. A similar program in Mexico, however, shows positive results on student test scores [[de Hoyos et al., 2017](#)]. In the net benefit we rank it at three widgets out of five, mainly because the cost of providing information is often next to zero and the benefit is potentially high (though uncertain). The difficulty of implementing its policies

is low and the expected time frame is medium-term.

4 Conclusion

The last nearly two decades have seen the WMS dataset expand by a factor of twenty, and the number of researchers using this data is over 5,000.²⁰ The first decade was primarily spent honing the measurement methods and amassing an arsenal of data with a focus on showing the various facets of the management and productivity relationship. In the second decade we started to see papers focusing on outcomes other than productivity, including energy efficiency, IT use and work-life balance. At the end of this second decade, the outcomes have expanded even further. Papers now in circulation explore the relationship between management practices and wage inequality, discrimination, tax avoidance, and more.

We also see an important set of papers pushing further on this idea of the manager vs. management, and the WMS very specifically focuses on the “technology” of it absent the manager. But the manager is the often the implementing party, and especially as hard technology starts changing, the job of the manager and the role of these partially office-based practices is also going to change. And finally, there is a question also on understanding the lack of adoption and to the extent that many of these practices are beneficial, why so many organizations fail to adopt them.

In terms of policy work, much has been done on structural drivers such as product market competition, foreign direct investment and regulation (even if the policy here is hard to do), but there is still much room for work on the “soft” type of drivers [Azulai et al.] and considering the effect of training, education, incentives in public vs private sector. Going into this more micro-level setting will take a large set of projects across various contexts to help us understand the frictions in each context. This is a fruitful area of research that we hope will continue to grow and develop.

²⁰A list of the over 170 projects approved for special data access are listed on our website, and the public anonymized dataset continues to be accessible to all interested researchers: www.worldmanagementsurvey.org

5 Tables and Figures

Table 1: Factor Analysis: management questions

	Manufacturing		Public hospitals			Public schools	
	<i>Factor 1</i>	<i>Factor 2</i>	<i>Factor 1</i>	<i>Factor 2</i>	<i>Factor 3</i>	<i>Factor 1</i>	<i>Factor 2</i>
Overall operations							
Adoption of lean	0.7521	-0.2944	0.4716	-0.4525	0.4387		
Rationale for lean	0.6898	-0.2883	0.5657	-0.4143	0.309		
Hospital standardization			0.6166	-0.2356	0.1513		
Hospital use of human resources			0.6145	-0.227	0.1453		
Standard instructional process						0.6089	-0.3457
Personalization of learning						0.7544	-0.1454
Data-driven planning						0.7402	-0.0209
Educational best practices						0.7269	-0.1527
Continuous improvement	0.7699	-0.2287	0.6755	-0.2262	0.0579	0.7411	-0.1582
Performance tracking	0.7691	-0.2664	0.6942	-0.0749	-0.2772	0.7433	-0.2851
Review of performance	0.7845	-0.2199	0.7361	-0.14	-0.3428	0.7525	-0.2858
Performance dialogue	0.7904	-0.2087	0.7287	-0.0481	-0.3066	0.781	-0.2216
Consequence management	0.7423	-0.0765	0.6823	-0.0487	-0.375	0.7286	-0.052
Type of targets	0.7151	-0.0634	0.6842	0.0746	-0.1241	0.7418	-0.1326
Interconnection of goals	0.7641	-0.0487	0.6848	-0.0496	-0.1569	0.7589	0.0166
Time horizon of goals	0.7021	-0.0517	0.63	0.0178	-0.0524	0.7564	-0.0509
Goals are stretching	0.6637	-0.0098	0.6771	-0.0253	-0.0949	0.7606	-0.0344
Clarity of goals	0.601	0.1125	0.5889	0.1219	0.0242	0.7355	0.1542
People management							
Talent mindset	0.656	0.2759	0.5892	0.2843	0.2821	0.6783	0.3687
Incentives and appraisals	0.5767	0.4075	0.5001	0.4784	0.0694	0.5558	0.274
Dealing with poor performers	0.4978	0.3606	0.547	0.3591	0.0614	0.6049	0.2954
Developing good performers	0.6525	0.3765	0.6325	0.3448	0.176	0.6477	0.4129
Employee value proposition	0.6684	0.2953	0.4792	0.3976	0.2591	0.4945	0.5207
Retaining talent	0.4413	0.4867	0.6301	0.0222	0.1999	0.7418	0.1791
Eigenvalue	8.48471	1.2507	7.83918	1.30441	1.0429	9.99207	1.22163
% total variance	0.4714	0.0695	0.392	0.0652	0.0521	0.4996	0.0611
Number of observations	12,209	12,209	1,125	1,125	1,125	1,326	1,326

Note: This table reports the results of conducting factor analysis on the World Management Survey sample for each of manufacturing firms, public hospitals and public schools. All questions from the three respective surveys are listed on the first column, and where the cell is blank the question does not exist for that particular survey. The questions from continuous improvement to retaining talent are common across all surveys. The questions of adoption of lean and rationale for lean are “common” across manufacturing and hospitals, though the scoring for each is sector-specific. Only the first two factors are listed because, for all three sectors, the third-plus factors had eigenvalues lower than one.

Table 2: Management Policy Toolkit

Policy type	Strength of evidence	Policy Net benefit (out of 5)	Difficulty of implementation	Time frame
Structural				
Competition	H	⊗⊗⊗⊗⊗	M	medium
Trade and FDI	H	⊗⊗⊗⊗⊗	L	medium
Education	M	⊗⊗	M	long
Labour Deregulation	M	⊗⊗⊗	L	medium
Governance	M	⊗⊗⊗⊗	M/L	long
Direct				
Training - consulting	H	⊗⊗⊗	H	short
Training - formal classroom	M	⊗⊗	H	medium
Information/benchmarking	L/M	⊗⊗⊗	H	medium

Note: This table is the summary of our (highly subjective) judgements of the literature on a core set of common policies. **Column 1** splits policies into “structural” (policies that not usually specifically targeted on management, but where research has produced evidence suggesting an indirect effect on management) and “direct” (policies more clearly targeted on management). **Column 2** includes our judgement of the strength of the evidence available for the effectiveness of each type of policy. Our decision of awarding low, medium or high is based on a combination of quantity, quality and credibility of research papers and research designs. Broadly, a ranking of *low (L)* refers to a body of evidence that primarily constitutes of cross-sectional correlational evidence. A ranking of *medium (M)* indicates that the evidence includes research designs with panel data, showing differences across time and some accounting of organizational fixed effects. A ranking of *high (H)* indicates that the evidence is arguably causal, with research designs including natural experiments or randomized control trials. **Column 3** scores the net benefit of the policy, taking into account the overall benefits *for firms* minus the costs incurred by the government, in widgets out of five. **Column 4** considers the difficulty of implementing these policies. While some might have high net benefits, they might be “costly” in terms of political capital and thus quite difficult to implement. Our implementation difficulty assessment is based on a combination of the likelihood of push-back from interest groups or policy recipients, the financial cost of implementing a policy, and the coordination requirement to get a policy enacted (including cross-departments or across countries). **Column 5** considers the time frame of when the expected effects would start materializing *after the policy has been enacted and fully implemented*. A short time frame refers to policies that would yield results within one year, a medium time frame refers to results between one and five years, and a long time frame refers to longer than five years.

Table 3: World Management Survey Questions: Core Operations

Question topic	Information collected
Manufacturing	
Adoption of modern practices	What aspects of manufacturing have been formally introduced, including just-in-time delivery from suppliers, automation, flexible manpower, support systems, attitudes, and behavior?
Rationale for adoption)	Were modern manufacturing techniques adopted just because others were using them, or are they linked to meeting business objectives like reducing costs and improving quality?
Hospitals	
Adoption of modern practices	What is the typical patient journey (or flow) through the hospital? How closely located are wards, theatres, diagnostic centers and consumables? How often do you run into problems with the current layout and pathway management?
Rationale for adoption	What was the rationale for improving the patient pathway? How often do you challenge/streamline the pathway? What factors led to the adoption of these practices?
Standardization of processes	How standardized are the main clinical processes? How clear are they to staff? What tools and resources do staff regularly employ? How do managers monitor protocol adherence?
Good use of human resources	What happens when different areas become busier than others? How do you know which tasks are best suited to different staff? What kind of procedures do you have to assist staff flow and coordination?
Schools	
Data driven planning and transitions	How is data used to inform planning and student transitions? What drove the move towards more data-driven planning and tracking?
Standardization of processes	How standardized are the instructional planning processes in the school? What tools and resources do teachers use to ensure consistent quality? How do leaders monitor and ensure consistency of quality cross classrooms?
Personalization of instruction	How much does the school identify and accommodate individual student needs? How do leaders ensure teachers are effective in personalizing instruction within classrooms? How are parents and students engaged?
Instructional best practices	How do leaders and teachers learn about instructional best practices? How do leaders encourage adoption and knowledge sharing across teachers? How do leaders ensure new practices are being used?

Note: This table lists the **core operations** topics covered in the World Management Survey questionnaires. The column “question topic” outlines the broad topic being measured. The last column includes a more detailed explanation of the types of follow-up questions that are asked of the manager to garner the information required for scoring. The set of questions on core/lean operations are the only set of questions that are industry-specific.

Table 4: World Management Survey Questions: Monitoring and target-setting

Question topic	Information collected
Manufacturing, hospitals and schools	
Process problem documentation	Are process improvements made only when problems arise, or are they actively sought out for continuous improvement as part of normal day-to-day processes?
Performance tracking	Is tracking ad-hoc and incomplete, or is performance continually tracked and communicated to all staff?
Performance review	Is performance reviewed infrequently and only on a success/failure scale, or is performance reviewed continually with an expectation of continuous improvement?
Performance dialogue	In review/performance conversations, to what extent are the purpose, data, agenda, and follow-up steps (like coaching) clear to all parties?
Consequence management	To what extent does failure to achieve agreed objectives carry consequences, which can include retraining or reassignment to other jobs?
Target balance	What type of goals does the organization have? Are they uni-dimensional (say, only financial for firms, or only government-assigned for public sector)? Is there a balance of targets?
Target interconnection	Are goals based on “shareholder value”? Are goals cascaded down the organization in a way that works through units and ultimately is connected to individual performance expectations?
Target time horizon	Do leaders focus mainly on the short term, or do they understand short-term targets as a “staircase” toward the main focus on long-term goals?
Target stretching	Are goals too easy to achieve, especially for some “protected/special” areas of the organization, or are goals demanding but attainable for all areas?
Performance clarity	Are performance measures ill-defined, poorly understood, and private, or are they well-defined, clearly communicated, and made public?

Note: This table lists the **monitoring and target-setting** management topics covered in the World Management Survey questionnaires. The column “question topic” outlines the broad topic being measured. The last column includes a more detailed explanation of the types of follow-up questions that are asked of the manager to garner the information required for scoring. This set of questions are common to all industries.

Table 5: World Management Survey Questions: People and incentives

Question topic	Information collected
Manufacturing, hospitals and schools	
Managing human capital	How do leaders show that attracting talent is important for the organization? To what extent are senior managers evaluated and held accountable for attracting, retaining, and developing talent throughout the organization?
Rewarding high performance	To what extent are people in the organization rewarded equally irrespective of performance level, or is performance clearly related to accountability and rewards?
Fixing poor performers	Are poor performers rarely removed, or are they retrained and/or moved into different roles or out of the company as soon as the weakness is identified?
Promoting high performers	Are people promoted mainly on the basis of tenure, or does the organization actively identify, develop, and promote its top performers?
Attracting human capital	Do competitors offer stronger reasons for talented people to join their organizations, or does a firm provide a wide range of reasons to encourage talented people to join?
Retaining human capital	Does the organization do relatively little to retain top talent, or does it do whatever it takes to retain top talent when they look likely to leave?

Note: This table lists the **people** management topics covered in the World Management Survey questionnaires. The column “question topic” outlines the broad topic being measured. The last column includes a more detailed explanation of the types of follow-up questions that are asked of the manager to garner the information required for scoring. This set of questions are common to all industries.

Table 6: World Management Survey Questions: Leadership

Question topic	Information collected
Hospitals	
Clearly defined accountability	What is the role of clinicians in improving performance and achieving targets? How are individual clinicians responsible for delivering targets? Does this apply to cost targets as well as quality targets? How do clinicians take on roles to deliver improvements?
Schools	
Leadership vision	What is the school’s vision for the next five years? Do teachers/staff know and understand the vision? Who are they stakeholders for the school and how is the vision communicated? How are they engaged in setting the vision?
Clearly defined accountability	Who is accountable for delivering on school targets? How are individual school leaders held responsible for targets? What authority do you have to impact factors that are important for delivering on the targets?
Clearly defined roles	How are the responsibilities of the school leader defined? How are responsibilities distributed across teachers and staff? How are roles defined? How are they linked to student performance?

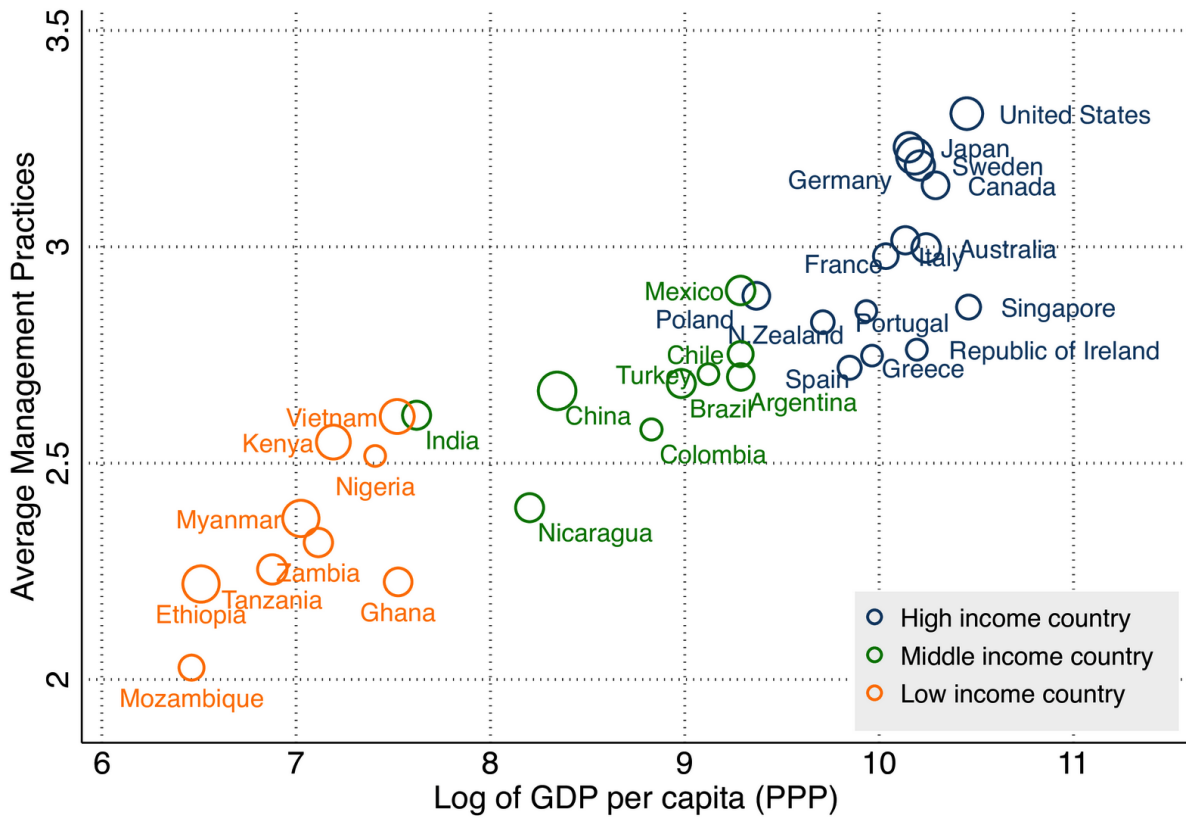
Note: This table lists the **leadership** topics covered in the World Management Survey questionnaires for schools and hospitals. These questions are not part of the manufacturing surveys. The column “question topic” outlines the broad topic being measured. The last column includes a more detailed explanation of the types of follow-up questions that are asked of the manager to garner the information required for scoring.

Figure 1: Measuring management: a summary of WMS-related methodologies

	Original methodology	Census-style	Adaptation of existing surveys
Features	<ul style="list-style-type: none"> • Phone-based • Trained analysts, 1hr interview 	<ul style="list-style-type: none"> • Self-respondent questionnaire • Sometimes mandatory 	<ul style="list-style-type: none"> • Relies on existing questions in large scale questionnaires • Primarily education ones
Pros	<ul style="list-style-type: none"> • Feasible for almost anyone • Higher response rates • (relatively) minimal infrastructure needed • Easy to calibrate • Less measurement error 	<ul style="list-style-type: none"> • Possibly cheaper • Potential of mandatory response (if census) • Potential of large N • Easier to do panels 	<ul style="list-style-type: none"> • Effectively free • Possibility of linking to other outcome data
Cons	<ul style="list-style-type: none"> • Very costly • Limits sample to those with phones • Logistical nightmare • 10-week maximum for projects (interviewer fatigue) 	<ul style="list-style-type: none"> • Potential of low response rate (if not mandatory) • Hard to compare across countries/languages • More prone to measurement error 	<ul style="list-style-type: none"> • Questions not always well-crafted • Questions dropped across years • Prone to measurement error

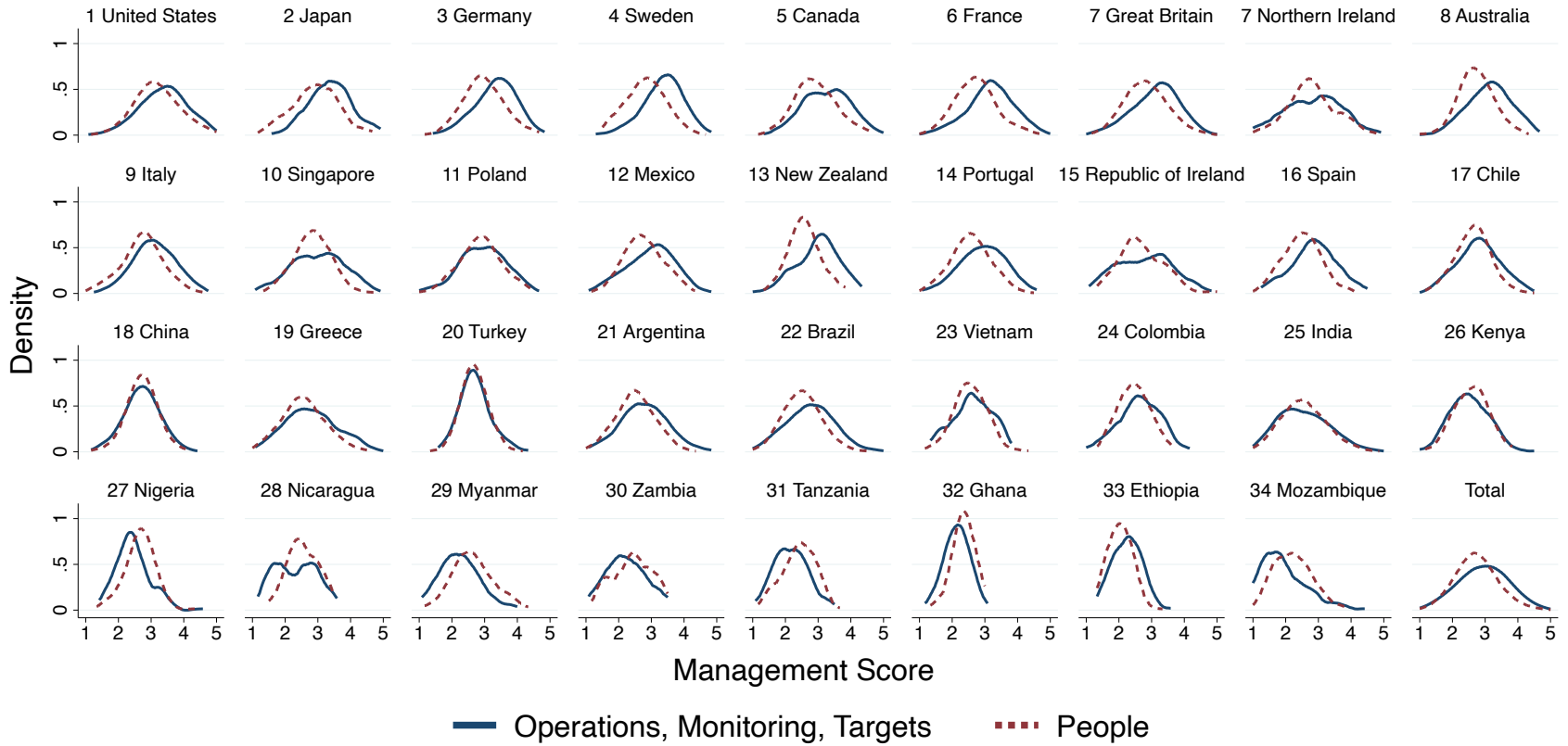
Note: This figure summarizes the main features, the pros and cons of the three primary styles of surveys that use (or are inspired by) the World Management Survey methodology.

Figure 2: Management and GDP per capita: manufacturing



Note: The y-axis is the country average management score from the World Management Survey (2004-2014). The x-axis is the log of the 2003-2013 average GDP per capita based on PPP from the IMF World Economic Outlook tables. Circle sizes represent median firm size.

Figure 3: Distribution of management across countries

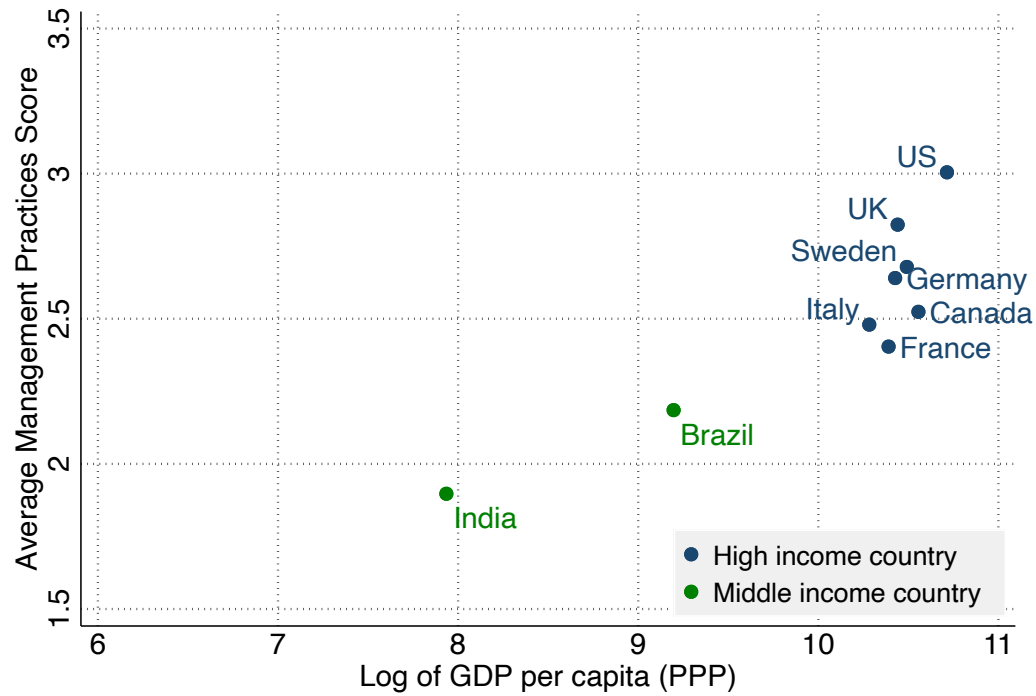


25

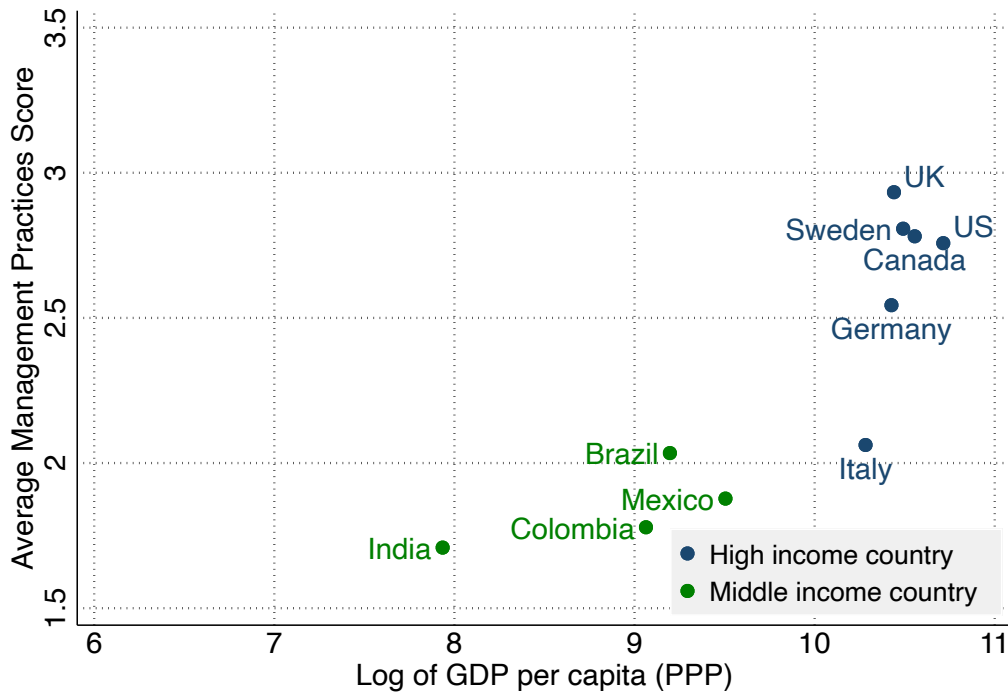
Note: This figure shows the distribution of the management scores across the 35 countries in the main 2004-2014 sample of the World Management Survey (2004-2014). The solid line represents the average operations management scores (including 12 questions on lean operations, monitoring and target setting) and people management scores (including the 6 questions related to incentives and personnel management).

Figure 4: Management and GDP per capita: public sector

(a) Hospitals



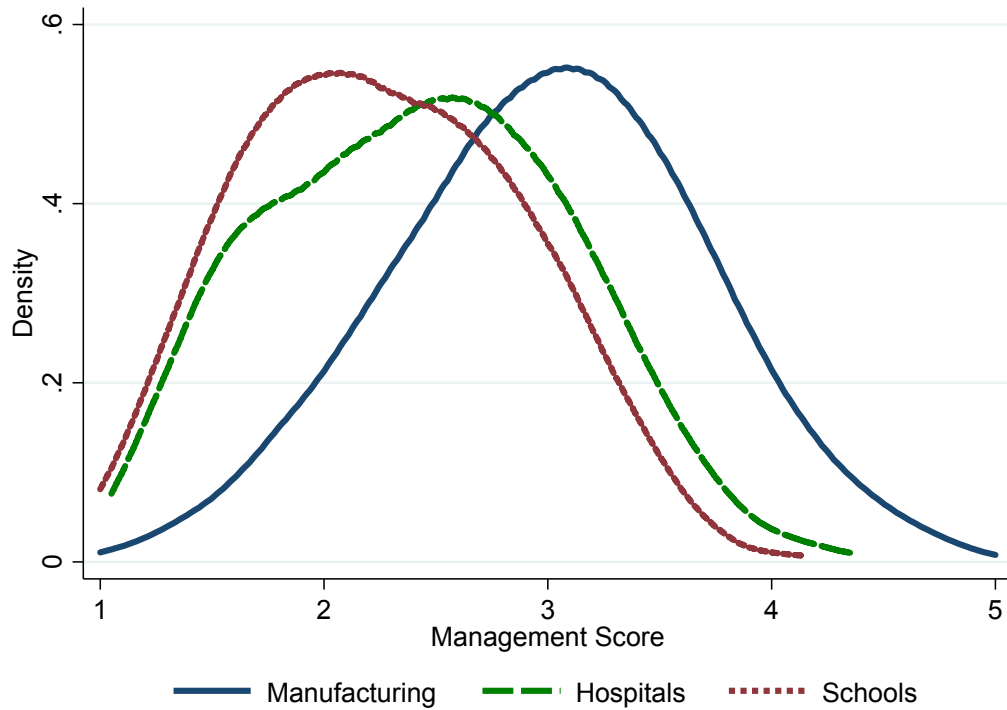
(b) Schools



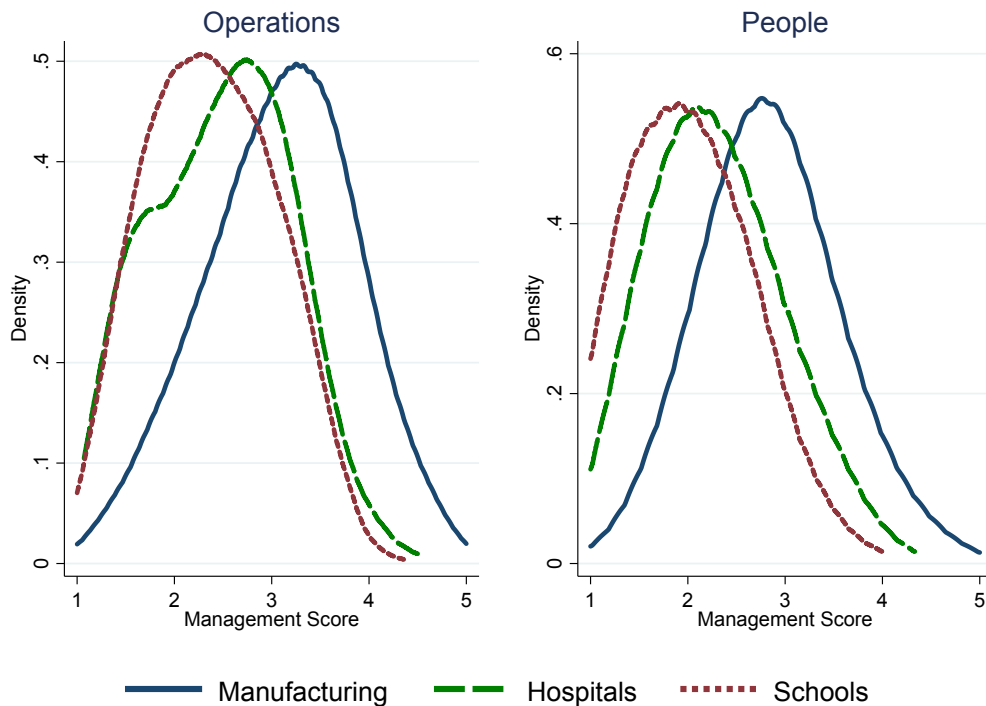
Note: The y-axis is the country average management score from the World Management Survey. Panel (a) includes scores from the WMS Hospitals survey, carried out in 2008-2009 for the US, Canada, UK, Sweden, Germany, Italy and France; 2012 for India and 2013 for Brazil. Panel (b) includes scores from the WMS Education survey, carried out in 2008-2010 for the US, Canada, UK, Sweden, Germany and Italy; 2012 for India; 2013 for Brazil; 2014 for Colombia and Mexico. Data for Colombia collected with Artur Harker, see [Bermudez and Harker \[2016\]](#). Data for Mexico collected with Rafael de Hoyos. The x-axis is the log of the 2003-2013 average GDP per capita based on PPP from the IMF World Economic Outlook tables.

Figure 5: Distribution of management across sectors

(a) Overall management score

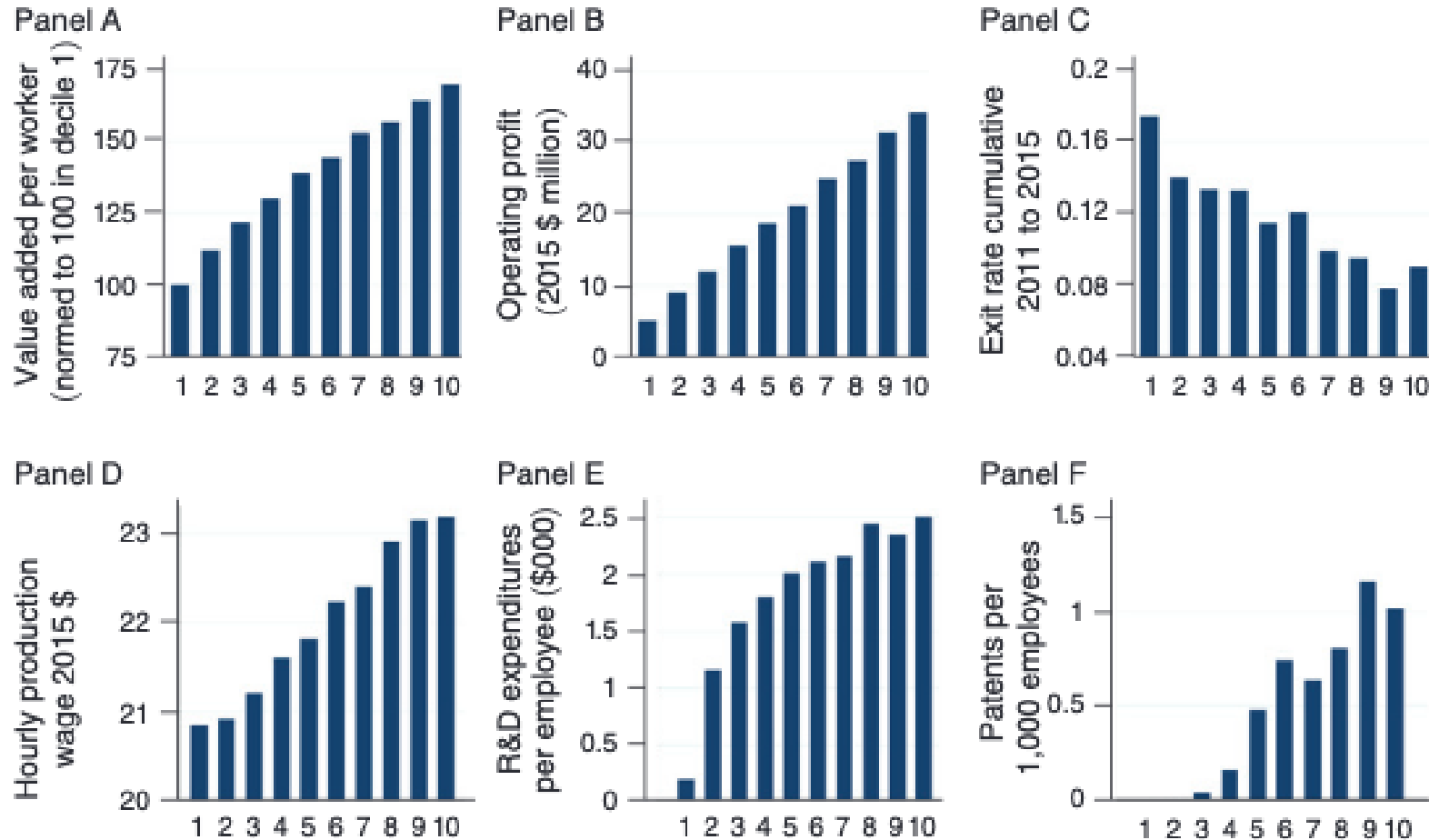


(b) Operations and people scores



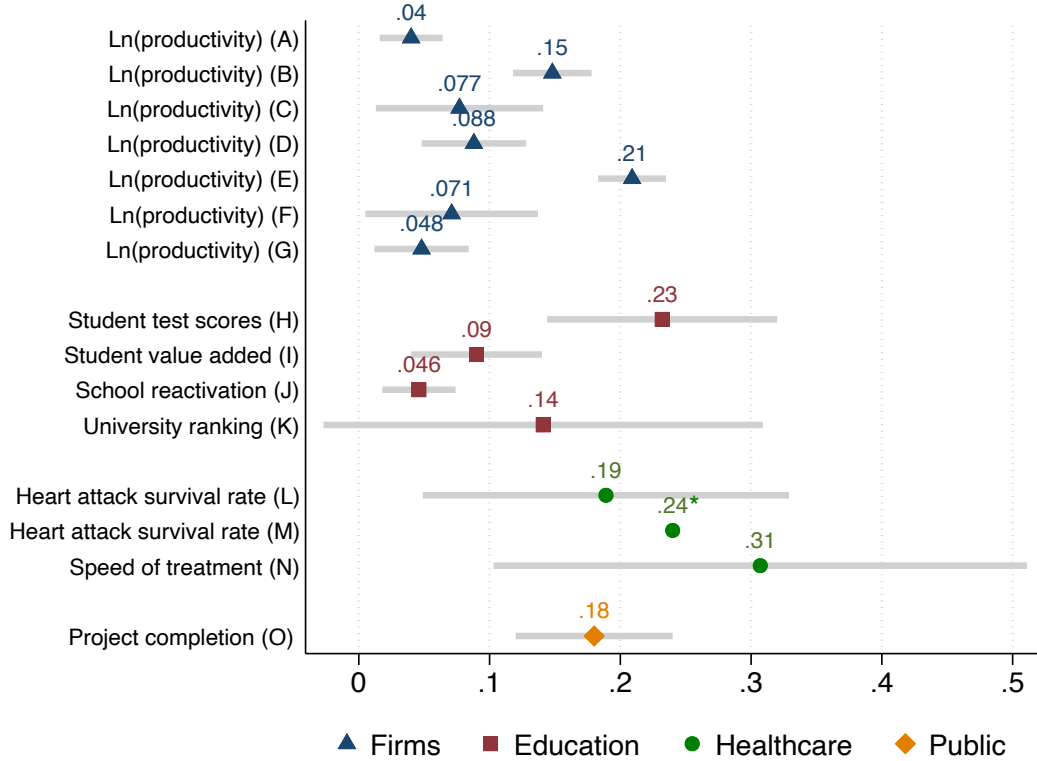
Note: This figure shows the distribution of the management scores across the 8 countries for which the World Management Survey has data across the three sectors: Brazil, Canada, Germany, India, Italy, Sweden, UK and US. Panel (a) plots the distribution for the overall management score for the set of questions that are common across sectors (questions on monitoring, target-setting and people management). Panel (b) plots the distribution of the operations management score (including questions on monitoring and target-setting) and people management scores. The solid lines represent the sample of manufacturing firms, the dashed lines represent the sample of hospitals and the dotted lines represent the sample of schools.

Figure 6: Management and performance



Source: Figure replicated from Bloom et al. [2019a], Figure 2. Notes from original figure: “The management score is the unweighted average of the score for each of the 16 questions, where each question is first normalized to be on a 0–1 scale. The sample in panels A, B, and F is all [management and organizational practices survey, MOPS] observations with at least 10 non-missing responses to management questions and a successful match to [Annual Survey of Manufacturers, ASM], which were also included in ASM tabulations, have positive value added, positive employment, and positive imputed capital in the ASM. The sample in panels C, E, and F only uses observations where we have 2010 MOPS management data. ‘Exit rate’ indicates whether a plant alive in 2011 had exited the economy by 2015. In panels E and F, we also condition on non-missing RD or patents requests count in the BRDIS survey. Management deciles are recalculated for the different samples. The figures are unweighted.”

Figure 7: Management and organizational performance



Legend: details of papers listed

Paper	Coverage	Outcome	Reference	Specification
A	Cross-country mfg firms	Log of productivity	Bloom and Van Reenen [2007]	Table 1, Col 4
B	Cross-country mfg firms	Log of productivity	Bloom et al. [2016]	Table 3, Col 2
C	Cross-country family firms	Log of productivity	Lemos and Scur [2019]	Table 1C, Col 1
D	Brazilian mfg firms	Log of productivity	Cornwell et al. [2020]	Table B2, Col 3
E	US mfg firms	Log of productivity	Bloom et al. [2019a]	Table 1, Col 2
F	Swedish mfg firms	Log of productivity	Heyman et al. [2019]	Table 8, Col 2
G	German mfg firms	Log of productivity	Bender et al. [2018]	Table 4, Col 5
H	Cross-country high schools	Student test scores	Bloom et al. [2015b]	Table 2, Col 3
I	India primary schools	Student value added	Lemos et al. [2021]	Table 5B, Col 1
J	Haiti schools	School reactivation	Adelman et al. [2020]	Table 2, Col 5
K	UK universities	University ranking	McCormack et al. [2014]	Table 6, Col 3
L	Cross-country hospitals	Survival (heart attack)**	Bloom et al. [2020a]	Table 2, Col 4
M	US hospitals	Survival (heart attack)**	Chandra et al. [2016]	Table 3, AMI
N	US addiction centers	Speed of treatment**	McConnell et al. [2009]	Table 3, Col 1
O	Nigerian civil service	Project completion rate	Rasul and Rogger [2018]	Table 4, Col 4

Note: This is a non-exhaustive list of papers that have correlated management with a productivity outcome for each type of organization. It is meant as a snapshot example of papers that use the World Management Survey methods (or similar) to measure management practices across industries and contexts. ‘Mfg’ is short for manufacturing. Productivity outcomes are coefficients that have sales or value added as the dependent variable but control for factor inputs (at a minimum capital and labor), such that the interpretation is the relationship between management and total factor productivity (TFP). Note that to improve readability we multiplied three of the coefficients by -1 such that all coefficients presented here have a “positive is better” interpretation. These are marked with ** on the list: the original papers that have coefficients where negative means a better outcome are: (L) and (M) risk-adjusted heart attack mortality rate, (O) days to treatment. The value for paper (M), marked with a * on the plot, does not have confidence intervals because it is a pair-wise correlation between the heart attack mortality rate. For paper (O), the coefficient refers to managerial *autonomy* and not a management score.

References

- M. Adelman, J. Barón, and R. Lemos. Managing shocks in education: Evidence from hurricane matthew in haiti. Mimeo, 2020.
- A. J. Ali, J. Fuenzalida, M. Gomez, and M. J. Williams. Four lenses on people management in the public sector: An evidence review and synthesis. *Oxford Review of Economic Policy*, forthcoming, 2021.
- D. Altig, J. M. Barrero, N. Bloom, S. J. Davis, B. H. Meyer, and N. Parker. Surveying business uncertainty. *Journal of Econometrics*, 2020.
- M. L. Anderson. Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American Statistical Association*, 103(484):1481–1495, 2008.
- D. Atkin, A. K. Khandelwal, and A. Osman. Exporting and Firm Performance: Evidence from a Randomized Experiment. *The Quarterly Journal of Economics*, 132(2):551–615, 02 2017.
- M. Azulai, I. Rasul, D. Rogger, and M. J. Williams. Can training improve organizational culture? experimental evidence from ghana’s civil service.
- O. Bandiera, L. Guiso, A. Prat, and R. Sadun. Matching firms, managers, and incentives. *Journal of Labor Economics*, 33(3):623–681, 2015.
- A. W. Bartik, M. Bertrand, Z. Cullen, E. L. Glaeser, M. Luca, and C. Stanton. The impact of covid-19 on small business outcomes and expectations. *Proceedings of the National Academy of Sciences*, 117(30):17656–17666, 2020.
- I. Ben-David, J. R. Graham, and C. R. Harvey. Managerial Miscalibration. *The Quarterly Journal of Economics*, 128(4):1547–1584, 09 2013.
- S. Bender, N. Bloom, D. Card, J. Van Reenen, and S. Wolter. Management practices, workforce selection, and productivity. *Journal of Labor Economics*, 36(S1):S371–S409, 2018.
- N. Bermudez and A. Harker. Factors associated with the quality of school management practices: an empirical analysis for colombia. Documentos de trabajo egob, universidad de los andes, 2016.
- K. Bilicka and D. Scur. Organizational capacity and firm profitability. Mimeo, 2020.
- N. Bloom and J. Van Reenen. Measuring and Explaining Management Practices Across Firms and Countries. *The Quarterly Journal of Economics*, 122(4):1351–1408, 11 2007.
- N. Bloom, T. Kretschmer, and J. V. Reenen. *Work-Life Balance, Management Practices and Productivity*, pages 15–54. University of Chicago Press, September 2009.

- N. Bloom, C. Genakos, R. Martin, and R. Sadun. Modern Management: Good for the Environment or Just Hot Air? *The Economic Journal*, 120(544):551–572, 05 2010.
- N. Bloom, C. Genakos, R. Sadun, and J. Van Reenen. Management practices across firms and countries. *Academy of Management Perspectives*, 26(1):12–33, 2012a.
- N. Bloom, R. Sadun, and J. Van Reenen. Americans do it better: Us multinationals and the productivity miracle. *American Economic Review*, 102(1):167–201, February 2012b.
- N. Bloom, H. Schweiger, and J. Van Reenen. The land that lean manufacturing forgot? *Economics of Transition and Institutional Change*, 20(4):593–635, 2012c.
- N. Bloom, B. Eifert, A. Mahajan, D. McKenzie, and J. Roberts. Does Management Matter? Evidence from India. *The Quarterly Journal of Economics*, 128(1):1–51, 2013.
- N. Bloom, M. Draca, and J. Van Reenen. Trade Induced Technical Change? The Impact of Chinese Imports on Innovation, IT and Productivity. *The Review of Economic Studies*, 83(1):87–117, 09 2015a.
- N. Bloom, R. Lemos, R. Sadun, and J. Van Reenen. Does management matter in schools? *The Economic Journal*, 125(584):647–674, 2015b.
- N. Bloom, C. Propper, S. Seiler, and J. Van Reenen. The Impact of Competition on Management Quality: Evidence from Public Hospitals. *The Review of Economic Studies*, 82(2):457–489, 01 2015c.
- N. Bloom, R. Sadun, and J. Van Reenen. Do private equity owned firms have better management practices? *American Economic Review*, 105(5):442–46, May 2015d.
- N. Bloom, R. Sadun, and J. Van Reenen. Management as a technology? Working Paper 22327, National Bureau of Economic Research, June 2016.
- N. Bloom, E. Brynjolfsson, L. Foster, R. Jarmin, M. Patnaik, I. Saporta-Eksten, and J. Van Reenen. What drives differences in management practices? *American Economic Review*, 109(5):1648–83, May 2019a.
- N. Bloom, P. Bunn, S. Chen, P. Mizen, P. Smietanka, and G. Thwaites. The Impact of Brexit on UK Firms. NBER Working Papers 26218, National Bureau of Economic Research, Inc, Sept. 2019b.
- N. Bloom, R. Lemos, R. Sadun, and J. Van Reenen. Healthy business? managerial education and management in health care. *The Review of Economics and Statistics*, 102(3):506–517, 2020a.
- N. Bloom, A. Mahajan, D. McKenzie, and J. Roberts. Do management interventions last? evidence from india. *American Economic Journal: Applied Economics*, 12(2):198–219, April 2020b.

- N. Bloom, K. Manova, J. Van Reenen, S. T. Sun, and Z. Yu. Trade and management. *The Review of Economics and Statistics*, 0(ja):1–44, 2020c.
- W. Brooks, K. Donovan, and T. R. Johnson. Mentors or teachers? microenterprise training in kenya. *American Economic Journal: Applied Economics*, 10(4):196–221, October 2018.
- T. Brown. Design thinking. *Harvard business review*, 86:84–92, 141, 07 2008.
- M. Bruhn, D. Karlan, and A. Schoar. The impact of consulting services on small and medium enterprises: Evidence from a randomized trial in mexico. *Journal of Political Economy*, 126(2): 635–687, 2018.
- J. Cai and A. Szeidl. Interfirm Relationships and Business Performance. *The Quarterly Journal of Economics*, 133(3):1229–1282, 12 2017.
- J. Cai and S.-Y. Wang. Improving management through worker evaluations: Evidence from auto manufacturing. Working Paper 27680, National Bureau of Economic Research, August 2020.
- A. Chandra, A. Finkelstein, A. Sacarny, and C. Syverson. Health care exceptionalism? performance and allocation in the us health care sector. *American Economic Review*, 106(8):2110–44, August 2016.
- A. Chong, R. La Porta, F. Lopez-de Silanes, and A. Shleifer. Letter Grading Government Efficiency. *Journal of the European Economic Association*, 12(2):277–298, 04 2014.
- C. Cornwell, I. M. Schmutte, and D. Scur. Building a productive workforce: the role of structured management. *Management Science*, (forthcoming), 2020.
- R. de Hoyos, V. A. Garcia-Moreno, and H. A. Patrinos. The impact of an accountability intervention with diagnostic feedback: Evidence from Mexico. *Economics of Education Review*, 58(C):123–140, 2017.
- F. A. Dunsch, D. K. Evans, E. Eze-Ajoku, and M. Macis. Management, supervision, and health care: A field experiment. Working Paper 23749, National Bureau of Economic Research, August 2017.
- A. Feng and A. Valero. Skill-Biased Management: Evidence from Manufacturing Firms. *The Economic Journal*, 130(628):1057–1080, 01 2020.
- A. Finizia. Managers and productivity in the public sector. Mimeo, 2020.
- R. G. Fryer. Management and student achievement: Evidence from a randomized field experiment. Working Paper 23437, National Bureau of Economic Research, May 2017.

- J. E. Galdón-Sánchez and J. Schmitz, James A. Competitive pressure and labor productivity: World iron-ore markets in the 1980's. *American Economic Review*, 92(4):1222–1235, September 2002.
- M. Giorcelli. The long-term effects of management and technology transfers. *American Economic Review*, 109(1):121–52, January 2019.
- G. K. Gosnell, J. A. List, and R. D. Metcalfe. The impact of management practices on employee productivity: A field experiment with airline captains. *Journal of Political Economy*, 128(4): 1195–1233, 2020.
- F. Heyman, P.-J. Norbäck, and R. Hammarberg. Foreign direct investment, source country heterogeneity and management practices. *Economica*, 86(342):362–395, 2019.
- Y. Higuchi, V. H. Nam, and T. Sonobe. Sustained impacts of Kaizen training. *Journal of Economic Behavior & Organization*, 120(C):189–206, 2015.
- Y. Higuchi, E. P. Mhede, and T. Sonobe. Short- and medium-run impacts of management training: An experiment in tanzania. *World Development*, 114:220 – 236, 2019.
- L. Iacovone, W. F. Maloney, and D. J. McKenzie. Improving management with individual and group-based consulting: Results from a randomized experiment in colombia. Technical report, The World Bank, 2019.
- D. Karlan, R. Knight, and C. Udry. Consulting and capital experiments with microenterprise tailors in ghana. *Journal of Economic Behavior & Organization*, 118:281 – 302, 2015.
- C. Leaver, R. Lemos, and D. Scur. Measuring and Explaining Management in Schools: New Approaches Using Public Data. CEP Discussion Papers dp1656, Centre for Economic Performance, LSE, Oct. 2019.
- R. Lemos and D. Scur. The ties that bind: Implicit contracts and the adoption of management technology in the firm. Technical report, CEPR Discussion Paper, 2019.
- R. Lemos, K. Muralidharan, and D. Scur. Personnel management and school productivity: Evidence from india. Working Paper 28336, National Bureau of Economic Research, 2021.
- K. J. McConnell, K. A. Hoffman, A. Quanbeck, and D. McCarty. Management practices in substance abuse treatment programs. *Journal of Substance Abuse Treatment*, 37(1):79 – 89, 2009.
- J. McCormack, C. Propper, and S. Smith. Herding Cats? Management and University Performance. *The Economic Journal*, 124(578):F534–F564, 03 2014.

- D. McKenzie. Small business training to improve management practices in developing countries: Re-assessing the evidence for “training doesn’t work. *Oxford Review of Economic Policy*, forthcoming, 2021.
- D. McKenzie and C. Woodruff. Business practices in small firms in developing countries. *Management Science*, 63(9):2967–2981, 2017.
- K. J. Meagher. Evidence on the non-linear impact of management. Working paper, 2013.
- K. Muralidharan and A. Singh. Improving public sector management at scale? experimental evidence on school governance india. Working Paper 28129, National Bureau of Economic Research, November 2020.
- V. H. Nam, N. T. T. Anh, and D. Q. Hung. *Kaizen for Small- and Medium-Sized Enterprises in Vietnam*, pages 247–267. Springer Singapore, Singapore, 2020.
- S. L. Parente and E. C. Prescott. Monopoly rights: A barrier to riches. *American Economic Review*, 89(5):1216–1233, December 1999.
- F. Queiró. Entrepreneurial Human Capital and Firm Dynamics. GEE Papers 00116, Gabinete de Estratégia e Estudos, Ministério da Economia, Dec. 2018.
- I. Rasul and D. Rogger. Management of bureaucrats and public service delivery: Evidence from the nigerian civil service. *The Economic Journal*, 128(608):413–446, 2018.
- I. Rasul, D. Rogger, and M. Williams. Management, organizational performance, and task clarity: Evidence from ghana’s civil service. Working paper, November 2018.
- J. Song, D. J. Price, F. Guvenen, N. Bloom, and T. von Wachter. Firming Up Inequality. *The Quarterly Journal of Economics*, 134(1):1–50, 10 2018.
- C. Syverson. Market structure and productivity: A concrete example. *Journal of Political Economy*, 112(6):1181–1222, 2004.
- C. Syverson. What determines productivity? *Journal of Economic Literature*, 49(2):326–365, 2011.
- A. Valero. Education and management practices. *Oxford Review of Economic Policy*, forthcoming, 2021.
- J. van Reenen. Does competition raise productivity through improving management quality? *International Journal of Industrial Organization*, 29(3):306–316, 2011.
- E. A. Verhoogen. Trade, Quality Upgrading, and Wage Inequality in the Mexican Manufacturing Sector. *The Quarterly Journal of Economics*, 123(2):489–530, 05 2008.