How Does Working-Time Flexibility Affect Workers' Productivity in a Routine Job? Evidence from a Field Experiment

Marie Boltz[*], Bart Cockx[†], Ana Maria Diaz[‡] & Luz Magdalena Salas [§¶]

July 21, 2022

Abstract

We conducted an experiment in which we hired workers under different types of contracts to evaluate how flexible working time affects on-the-job productivity in a routine job. Our approach breaks down the global impact on productivity into sorting and behavioral effects. We find that flexible arrangements that allow workers to decide when to start and stop working increase global productivity by as much as 50 percent, 40 percent of which is induced by sorting, and 60 percent represents a motivational effect, mainly driven by more effective working time, with workers reducing the length of their breaks. Our findings also suggest that part-time contracts can enhance global productivity, – though not significant at conventional levels –, and that this effect is also driven by a significant drop in the length of breaks taken. We hence contribute to the literature providing causal evidence of flexibility in routine jobs leading to higher productivity.

JEL Codes: J21, J22, J23, J24, J33

Keywords: Flexible work arrangements, productivity, labor market flexibility, work–life balance, part-time work.

[*]BETA, University of Strasbourg, 61 avenue de la Forêt Noire, 67085 Strasbourg, France, m.boltz@unistra.fr

[†]Department of Economics, Ghent University; IZA, Bonn; CESifo, Munich; IRES, Université catholique de Louvain; ROA, Maastricht University, bart.cockx@ugent.be

[‡]Departamento de Economía, Pontificia Universidad Javeriana, a.diaze@javeriana.edu.co

[§]Departamento de Economía, Pontificia Universidad Javeriana, salas.luz@javeriana.edu.co

[¶]This paper has benefited from great research assistance by Juan David Hernandez and Juan Sebastian Ortiz. We thank Nicolas Lillo for his work building the software and providing the Chilean Census data. The authors would also like to thank Pontificia Universidad Javeriana, University Paris 1 Panthéon Sorbonne, and the University of Ghent for financial support. Ethics Statement: The research was performed on human subjects and received IRB approval from Universidad Javeriana on October 3, 2018.

Introduction

Flexibility in working arrangements[1] – part-time work, flexible time schedules, or teleworking – are typically either chosen by the employee – often as a way to achieve a better work-life balance (OECD, 2016) – or imposed by the employer (Schneider and Harknett, 2019). While we see an increasing trend in the use of these flexible arrangements, we still know little about how they impact on-the-job productivity and self-selection into flexible jobs. In this paper, we explore in an experimental setting whether offering flexible working arrangements to new hires has an impact on productivity for routine jobs – separating the selection effect from the on-the-job motivational effect. We focus on two types of flexibility: part-time, and flexible time schedules.

Working-time and workplace flexibility play an essential role in promoting an adequate work–life balance, especially for working parents (OECD, 2016). Part-time work is an option for employees who want to reduce their working time, but it comes at a price – reduced earnings.[2] Work-schedule flexibility, in which workers can determine when to start and finish work each day, does not generally entail such an earnings penalty.[3] In Europe, about one-third of employees report having some form of control over their working time; far fewer employees in routine occupations have this option (Eurofound, 2015). This figure is similar in the US, and it depends on the type of job and the characteristics of the worker (Golden, 2001).

A key question regarding such flexible working arrangements is why more firms do not offer them. If workers value such arrangements, they are willing to work for lower

[1] Spreitzer et al. (2017) provide different definitions of flexibility in working arrangements.

[2] In 2019, 9.6 percent of male employment and 25.4 percent of female employment in Organisation for Economic Co-operation and Development (OECD) countries were part time (OECD, 2020). In Colombia, where our study was conducted, the figures are very similar: 9.0 percent for men and 26.5 percent for women, which also tracks closely with the rates from EU28 countries: 8.0 percent for men and 26.3 percent for women.

[3] Some studies show, however, that flexible work practices can entail career advancement penalties (Leslie et al., 2012), and especially so for women (Brescoll et al., 2013).

wages in exchange for such flexibility (Bustelo et al., 2022; Chen et al., 2020, 2019; He et al., 2021; Mas and Pallais, 2017). Alternatively, when keeping pay constant, the option of such autonomy may boost a worker's productivity either by increasing their motivation or stimulating reciprocal behavior (Beckmann et al., 2017; Rupietta and Beckmann, 2017). Currently, flexible working time is largely given to workers in higher-skilled occupations (Golden, 2001; Wiß, 2017). One key reason is that employers may worry that giving workers with routine jobs more freedom will cause them to work less or to reduce their level of effort, which comes at the expense of firms' performance (Beckmann, 2016; Ortega, 2009). A similar line of argument is observed for women who, in the European context, are found to face the double penalty of reduced access to flexible arrangements and lower wages in female-dominated workplaces (Chung, 2019; Magnusson, 2021). While the effect of workplace flexibility – teleworking – on productivity has been the object of considerable study, this is less the case for working-time flexibility, especially for routine jobs.

The objective of this paper is to investigate how productivity on-the-job differs across contract types when we vary the degree of working-time flexibility of the job. We also aim at disentangling whether productivity differences induced by the different contractual arrangements result from self-selection,[4] or from workers' motivation on-the-job. We therefore conducted a field experiment in Bogotá, Colombia's capital city, to provide causal evidence of the effects on productivity in a *routine* temporary job of two flexible working-time arrangements: (1) working part time and (2) being able to decide when to start and stop working within the work week.

[4]We use the terminology of self-selection in this paper to define the selection effect in the application stage of the workers: we test whether more productive workers – as measured by an *ex-ante* productivity test – are more or less likely to accept the job offered when the contract allows for more flexibility relative to a non-flexible contract. We are *not* considering a framework where people are offered a menu of contracts that differ in terms of flexibility, and where they have to sort themselves into one type of contract. Rather, we randomly allocated contract types across individuals and observed whether the level of *ex-ante* productivity varied across job types.

Flexible work schemes can affect productivity through different channels. First, they can increase the attractiveness of the job such that intrinsically more productive workers *self-select* into it. Second, they incentivize workers to exert more effort on-the-job or to spend more time working by increasing worker's motivation or reciprocity towards the employer (Beckmann, 2016; Kossek and Michel, 2011). Third, they can affect productivity negatively by increasing coordination costs (Gibbs et al., 2021) or by making shirking behaviors more difficult to monitor (Nagin et al., 2002). Our field experiment aims to test for these channels and to disentangle the global effect into these components.

The field experiment was designed as follows. We first posted a vacancy for a three-week position for data entry operators in a prestigious university. The contractual conditions were not mentioned in the vacancy. Subsequently, applicants were asked to perform an online test in which they had to carry out similar data entry tasks as the job to which they were applying. This enabled us to construct an *ex-ante* measure of productivity that we use to disentangle the self-selection effects on productivity from the on-the-job motivational effects. Subsequently, we randomly allocated the 535 applicants to one of four contract environments: (i) full-time non-flexible schedule, (ii) part-time non-flexible, (iii) full-time flexible, (iv) part-time flexible – where full-time refers to 40 hours per week, part-time to 20 hours, non-flexible means that hours of work were imposed, while the flexible schedule allowed workers to manage their working hours. Applicants were then informed of the assigned contractual environment, including the hourly wage, which was set at approximately 1.3 times the minimum wage. Of the candidates who were still interested, we then randomly offered positions for each contractual environment. We filled 38 positions, among which 34 workers stayed until the end of the contractual period. On the job, we then monitored the *overall (ex-post)* productivity of the hired workers by measuring the average number of correctly typed characters over the contractual time. We also measured the actual working time, the duration of breaks, and the periods of absenteeism. Based on these measures, we could break down global productivity into measures of speed, precision, effective working time, and work interruptions. This allowed us to perform an in-depth analysis of the mechanisms at play.

The point estimates suggest that an employer can increase a worker's overall productivity by nearly 50 percent by offering a full-time flexible contract rather than a full-time non-flexible one. About 40 percent of this overall effect is attributed to attracting more productive workers. The on-the-job motivational effect is almost completely driven by the fact that full-time flexible workers take fewer breaks than those with non-flexible contracts. This results in a 10-percentage-point increase in effectiveness relative to contractual working time. This is in line with the results of Beckmann et al. (2017), who, based on the German Socio-Economic Panel, also find that flexible work-time scheduling arrangements increase actual working hours relative to contractual working hours. Similarly, using observational and survey data from US insurance and telecommunications companies, Pierce and Newstrom (1983) and Kossek and Michel (2011) find that workers under more flexible working arrangements had significantly reduced absenteeism.

Our findings also suggest that part-time contracts can enhance global productivity, but these effects are not significant at conventional levels. However, we do find that part-time workers, with or without a flexible time arrangement, spend 15 percentage points less time in breaks relative to contractual time than the reference group, and this effect is highly significant. This does not show up in the global effect on productivity, because these part-time workers seem to make more mistakes, or to be more frequently absent from work.

Our finding that working-time flexibility can substantially enhance productivity in a routine data entry task is new. It has strong policy implications: we show that introducing more working-time flexibility could be Pareto improving for both employers and employees. This is even more important because lower-wage workers often do not have access to this type of flexibility and are more likely to benefit from it (Kim, 2020).

Prior studies have argued that flexibility is especially valuable in non-routine, "creative" tasks, because autonomy would matter more in these kinds of tasks than in "dull" routine tasks (Beckmann, 2016; Dutcher, 2012). A potential explanation for this is that the existing literature focused on the effects of working from home. A major distinguishing feature is that workers are not monitored when they work from home, while in our setting they are. This monitoring may reduce the incentive to shirk. An additional potential explanation is that the experiment was set up in an environment of a less developed country, where more than half of the workforce works in precarious informal jobs with remuneration below the minimum wage. In such an environment, it is more likely that workers who are attracted to a (temporary) routine job paying 30% more than the minimum wage would in the counterfactual work in informal jobs. Such workers may value flexibility a lot, and may prefer a full-time to a part-time job, because the opportunity cost is larger for the latter. This can explain why the full-time flex job attracts relatively more productive workers than the other contract types.

Another main contribution to the literature consists in studying the productivity effects of working-*time* flexibility, rather than of working-*place* flexibility, and in proposing a method to decompose the effect of working-time flexibility on global *ex-post* productivity into an *ex-ante* selection effect induced by attracting workers with intrinsically higher productivity levels, and an *ex-post* motivational effect on-the-job. To the best of our knowledge, Bloom et al. (2015), and Harrington and Emanuel (2020) are the only ones who have studied the selection effects of flexible working arrangements, but they focus solely on working remotely, and on workers who are already hired. We are the first to study this selection effect during the hiring process, and for working time flexibility.

The remainder of the paper is organized as follows. First, we summarize the literature on the effects of flexible working arrangements on productivity and other outcomes. Then, we describe the experimental design in more detail. Subsequently, we present the data and associated descriptive statistics, and define the outcome variables and their decomposition. Next, we assess how flexible working arrangements affect on-the-job productivity overall, including both selection and motivational effects. Then, we break down the overall effect to explore the mechanisms at work, and report the results of our robustness checks. The final section concludes.

Literature Review

The literature has mainly focused on the impact of flexible working arrangements on overall well-being and work–life balance, and less on productivity at work. Moreover, the focus was mostly on unpredictable work schedules imposed by the employer (e.g. Golden, 2015; Henly and Lambert, 2014; Kamalahmadi et al., 2021; Lambert, 2008; Schneider and Harknett, 2021), rather than on the freedom to choose working hours.[5] The few papers that explore the effect of working arrangements on productivity concentrate on working from home. Our paper contributes to this literature in that it explores how working-time flexibility affects firms' outcomes, and productivity in particular.

There are many measures of productivity and the choice among them often depends on data availability. Studies analyzing observational data use total hours worked, overtime, job satisfaction, or absenteeism as proxies for productivity. Most of these studies

[5]The study by Allen et al. (2013) provides an extensive review of the literature that establishes a relationship between different forms of flexibility and work–family conflict. The book by Kelliher and De Menezes (2019) provides a thorough discussion about definitions and motivations for flexible work, how flexible work has been introduced by policy makers in some countries, and how to assess flexible working arrangements on employers' outcomes – i.e. retention, performance, commitment – and employees' outcomes – i.e. well-being and job satisfaction. These latter outcomes have also been studied by Allen et al. (2013); Chung and Van der Horst (2018); Golden (2015); Golden et al. (2018); Henly and Lambert (2014); Kamalahmadi et al. (2021); Lambert (2008); Leslie et al. (2012); Schneider and Harknett (2021).

find a positive effect of flexible working arrangements on those measures of productivity. For example, Kelliher and Anderson (2010) and Eaton (2003) use original survey data from UK and US firms, respectively, and find that workers with more flexible schedules are more satisfied and committed at work, and thus exert more effort, possibly induced by reciprocal behavior. Beckmann et al. (2017) use data from the German Socio-Economic Panel (SOEP) to find that job flexibility induced an increase in total hours worked and a reduction in absenteeism, explained mainly by intrinsic motivation of workers and less by a reciprocal behavior. Similarly, also using the SOEP data, Rupietta and Beckmann (2017) find that intrinsic motivation serves as a mediator of the impact of workplace flexibility on effort. Employing a structural approach using matched employer–employee data for the Dutch pharmacy sector, Nelen et al. (2011) explore how firms offering part-time contracts in addition to full-time contracts are more productive – as measured by the number of prescription lines delivered to customers per day at the firm level – than firms with a large-share of full-time workers.

Some studies highlight more mixed results, depending on the form of flexibility or on the target population. For instance, based on the SOEP, Lott and Chung (2016) conclude that schedule control increases overtime hours and income, but only for men. In the UK, Chung and Van der Horst (2020) find that income and unpaid overtime working hours increase for workers who experience schedule control, but not for workers with family-friendly flexible working arrangements – i.e. chosen part-time and telework. Finally, more recently, in the wake of the pandemic, Harrington and Emanuel (2020) and Gibbs et al. (2021) study the causal effects of remote work on productivity in the quasi-experimental setting that COVID-19 generated, and provide evidence that working time flexibility is detrimental to productivity in some cases. Harrington and Emanuel (2020) find that remote work raises productivity by 7% to 10%. However, individuals who were newly hired during the pandemic and had to work remotely were 12% less productive than on-site workers. Similarly, Gibbs et al. (2021) find a 20 percent reduction in the overall productivity for those working from home as a result of an increase in the number of hours worked that did not translate into more output per worker.

Another strand of literature relies on field experiments to evaluate the effect of flexibility on workers' productivity. This helps to draw causal links concerning how flexible arrangements affect productivity and so to derive direct measures of productivity. Closely related to our work, four recent field experiments evaluated the effects of workplace flexibility on workers' productivity. First, Dutcher (2012) assigned students randomly to "creative" and "dull" tasks – data entry tasks as in our study. Each of these tasks was carried out either inside or outside a laboratory, respectively representing working from the office and from home. The author finds that telecommuting enhanced productivity in the creative task, but had negative effects in the dull one. The lack of monitoring of those who worked from home may explain this finding, because in a dull task the positive motivational effect of autonomy on productivity is much smaller than in a creative task, and therefore the risk of shirking in the absence of monitoring is higher (Beckmann, 2016). Second, Bloom et al. (2015) offered employees from a Chinese call center the option to work from home. Those interested were randomly offered positions to work from home for a 9-month period after which an increase in self-reported satisfaction and a productivity gain of about 13 percent was observed. A large part of the effect, namely 9 percent, was due to working longer hours, while the remaining 4 percent resulted from making more telephone calls per hour. In addition, the study measured sorting effects, but only among the already hired workers, and found that teleworking could further enhance the productivity gains to 22 percent. Third, Angelici and Profeta (2020) randomly offered a "smart work" alternative to a sample of over 300 workers from an Italian firm, and measured the effect on productivity, well-being, and work–life balance. Similar to ours, workers had the opportunity to choose the time of work, but also the place of work (outside the office), one day a week during a 9-month period. Offering ”smart working” increased self-reported and objective productivity, and improved well-being and work–life balance. Lastly, in the study by Williams et al. (2018) administrators from a large retail store offered workers more stable shifts and the possibility to swap shifts at their request, providing workers with some freedom to manage their own schedules, and it turned out to be very effective at increasing workers' productivity and store sales, by 5 and 7 percent, respectively.

Building on these experimental papers, our paper contributes to the literature by measuring the impact on productivity of offering new real short-term job contracts with working schedule flexibility by separating both selection in job acceptance and motivational effects on-the-job.

Experimental Design

We conducted a field experiment in which we posted real job ads for data entry clerks, so as to observe job applicants' choices and performance in a natural environment. We generated random variation across job seekers in invitations to apply for jobs that differed in two dimensions of job schedule flexibility: working time (full time versus part time) and, for any given working time, the autonomy to schedule when to accomplish this work within the workweek. This resulted in randomizing job seekers into one of four contract environments: non-flexible full time, flexible full time, non-flexible part time, and flexible part time.

The experiment was designed to measure the effect of job schedule flexibility on productivity. This flexibility can affect productivity in two ways. First, inherently more productive workers may be more likely to accept jobs that provide the flexibility they prefer. Second, conditional on being hired, workers may allocate more or less effort to the job. To measure the total effect of this *ex ante* self-selection and *ex post* motivational reaction to flexibility on workers' productivity (on-the-job), we advertised the position without referring to the contract environment and then randomly offered applicants to *one* of the four arrangements described above. After imposing certain prerequisites designed to mimic the recruitment selection process, a random sample of remaining candidates was offered the position. The workers who accepted the position were hired for a period of 3 weeks. Their on-the-job productivity was measured throughout this time.

Steps of the hiring process

Table 1 summarizes the sample size at each stage of the experiment, and thus sets out the recruitment process.

Step 1: The job ad

We placed real job advertisements for data entry clerks every week for a month in different standard job postings in Bogotá (internet and newspaper). The job offer involved real employment for a short period of time and a task that did not require a specific level of education or specialized skills. We advertised the position without referring to the type of contract, and asked job seekers to fill out an online CV form, which included an online productivity test. The posted job ad was phrased as follows:

Job Advertisement Title: Prestigious university needs data clerks to support a research project.

Description: Contract for the provision of services.

Duration: Three weeks.

If you are interested in this offer, apply via the following link: LINK, or send us a message via Whats-App.

We received 686 job applications to the job ad.

Step 2: Collecting resumé information and the productivity test

The online CV form included standard questions about the applicants' level of education, labor market experience, and other demographic characteristics (see the questionnaire in the supplemental material). After filling out the form, applicants were invited to complete an online test to measure their *ex ante* productivity levels.[6] Those who started the application but did not finish the online test were not considered in the next stages of the application process and were dropped from the experiment. The test consisted of executing three tasks that resembled those required in the job to which the job seeker applied.[7] Based on the answers to this test, we constructed the following *ex ante* productivity measure for each individual, i, and each task, T:

$$Prod_{iT} = \frac{Correct\ Answers_{iT}}{Total\ Questions_{iT}} \times \frac{1}{Time_{iT}}$$

An applicant was classified as a highly productive worker if she scored higher than the median in all three tests:

$$High_i \quad = 1[Prod_{i1} \succeq Med(Prod_{i1}) \& Prod_{i2} \succeq Med(Prod_{i2}) \& Prod_{i3} \succeq Med(Prod_{i3})] \quad (1)$$

[6]The tests were timed, but job applicants did not know this.
[7]See the description of the test in the supplemental material.

where $1[A] = 1$ if A is true, and $1[A] = 0$ if A is false.

Step 3: Random assignment of contract types

The 535 respondents who completed the required resumé information and the productivity test were randomly assigned to one of the four contract types: T1 — full-time non-flexible (control group), T2 — part-time non-flexible, T3 — full-time flexible, and T4 — part-time flexible. They were informed about the contractual terms of the job, the place of employment, wage, and working-time regime.[8] The first three contractual terms were fixed across all treatments: non-renewable 3 weeks contract between November 26 and December 14, 2018 in the computer labs of the university,[9] remunerated at a gross hourly wage of 7.000 Colombian Pesos (COP), which is about 2.33 USD. This is approximately 1.3 times the minimum wage or 74 percent of the average wage in Colombia. The dimensions that varied across contract types were the four working-time regimes.

The full-time contract required working 40 hours a week, and the part-time contract was for 20 hours. In a fixed working-time schedule the full-time worker had to work Monday to Friday, 8 AM to 5 PM with a one-hour lunch break from noon to 1 PM. Part-time workers were randomly assigned to work either in the morning (from 8 AM to 12 PM) or the afternoon (from 1 PM to 5 PM). In a flexible working-time schedule the worker could freely determine their working hours within the opening hours of the computer lab, i.e. Monday to Friday, 8 AM to 8 PM.

We stratified the sample before randomizing applicants to one of the four treatment conditions according to their *ex ante* productivity, because we expect to be a good predictor of the potential outcome, *ex post* or on-the-job productivity (Imbens and Rubin, 2015). Thus, we split the sample into high or low *ex ante* productivity as defined in equa-

[8]We present the phrasing of this e-mail in the supplemental material.
[9]We did not offer the possibility of teleworking.

tion (1) in step 1. Also, we used gender and being a caregiver as additional stratification variables since contractual flexibility is particularly relevant for women in general and caregivers in particular.[10] This resulted in eight stratification cells defined by all possible combinations of these three stratification variables.

Step 4: Mimicking the recruitment selection process

To mimic a typical recruitment process, we sent another e-mail asking additional open-ended questions about their motivation for the position to all those who responded affirmatively and those who did not respond to the previous e-mail (see the second email in the supplemental material). 384 individuals stated their interest in the position after this second email and after knowing their contract type.

In a fourth step, job offers were sent to a randomly chosen sub-sample of applicants who reacted positively to this last e-mail. For this purpose, of the 384 retained candidates, 13 were randomly selected from each contract type, so 52 (= 13 x 4) job offers were first sent out. However, not all of these offers were accepted, likely due to the standard additional administrative requests involved.[11] Consequently, each applicant who did not accept the job offer in a particular treatment condition was replaced by another one randomly drawn from the same stratification cell. Because we were constrained by the announced starting date of the employment contract, we did not manage to hire the intended 13 workers in all of the four contract types. Eventually, after offering a job to 79 applicants, we filled 38 positions; four of these quit shortly after the contract started.

Table 2 shows the distribution of offers and hired individuals across contract types.

[10] Caregiver is a person who has to take care of children under 5 years of age, a disabled person or an adult over 65 years old with permanent care needs.

[11] The administration requests the provision of a tax identification number, an identification document, a certificate of affiliation with a health service, and a declaration that one is willing to pay the mandatory social security contributions (see supplemental material for more details).

It shows that the standard schedule (i.e. full time, no flex) had a higher acceptance rate than the other options.

The work environment

The workers were hired to perform a data entry task in the computer labs at Universidad Javeriana in Bogotá. We assigned a separate lab to each treatment group, located in different buildings, to reduce the risk of communication between individuals from different treatment groups. We hired eight students to manage the labs who were randomly assigned to one of the four rooms in two shifts (morning and afternoon), and were rotated across all rooms to minimize assistant-specific effects. Assistants were present at the computer lab all the time: they did not monitor the workers, but they reported any activity that happened in the room. They also registered each worker's time of arrival and departure every working day, and assisted with any technical difficulties. Additionally, they opened the computer labs every morning and closed the rooms after everyone had left. They also instructed workers on how to use the software before starting the data entry tasks.

An ID code was assigned to each worker, together with a user name and password to log into the software created for the data entry process. The software displayed the image in a dialog box (typing space), recorded the answer, the time spent on each image, the breaks reported by the worker, and the time the workers completed the tasks for the day. This software was created in Python and depicted images of characters (numbers) from a Chilean Agrarian Census with 350,000 images.[12]

Each image was randomly assigned with equal probability to approximately nine workers to determine the correct data entries. As a result, on average all the images

[12] The data entries were used for a research project of professor Nicolas Lillo from the Economics Department at Universidad Javeriana.

from the Census were typed at least seven times by seven different workers, because not all workers typed at the same speed and therefore did not type all the planned images. Figure 1 shows an example of what was displayed on the screen.

Experimental sample and data

This section presents the measures of productivity used in the analysis and describes the characteristics of the different samples – from the pool of applicants to the final group of hired workers in our experiment.

The measures of productivity and their decomposition

Our main aim is to determine the impact of working-time flexibility on labor productivity. We construct a measure of *ex post* productivity of hired workers and we break it into different components to explore the mechanisms of potential productivity differences. We also decompose our measure of *ex ante* productivity into efficiency, speed, and precision, which we use to disentangle the selection from the *ex post* motivational effects.

Ex post on-the-job productivity

Different measures of productivity could be considered. In a private-sector environment, a natural measure of productivity would be the value of production per unit of contractual working time. In the current context, the value of production is not known, but since the data entries served as an input for research purposes it is natural to value accuracy. This was stressed when instructions were provided on the first working day. We therefore code each correctly typed character as 1 and each incorrectly typed character as 0 and measure the productivity level of worker i in time period t as:

$$AP_{it,0} \equiv C_{it}/T_{it} \tag{2}$$

where C_{it} is the number of correct characters that individual i typed within the contractual time T_{it} of period t. In the analysis below we consider that $t \in \{1,2\}$ for both working weeks included in the analysis. For a full-time worker $T_{it} = 40$ hours, while for a part-time worker $T_{it} = 20$ hours.

We determined a correctly typed character to be equal to the mode of all typed characters for each corresponding character image.[13] To have an homogeneous measure of productivity throughout the period, we focus only on the first 2 weeks of the contract, because in the last week we assigned a different task that was more complex and thus not comparable with the one performed during the first 2 weeks.[14]

To explore the mechanisms that generate differences in productivity between types of contracts, we define our productivity measure as follows:

$$AP_{it,0} \equiv C_{it}/T_{it} = (C_{it}/N_{it}) \times (N_{it}/D_{it}) \times (D_{it}/T_{it}) \equiv \prod_{j=1}^{3} AP_{it,j} \tag{3}$$

where N_{it} is the total number of characters typed within the contractual time T_{it} and D_{it} is the *actual* working time. The above equation breaks down average productivity ($AP_{it,0}$) into a measure of *precision* (C_{it}/N_{it}), i.e. the fraction of correctly typed images, a measure of *speed* (N_{it}/D_{it}), i.e. the number of images typed per unit of actual working

[13] In a sensitivity analysis we used the median instead of the mode, but this did not change our findings. This analysis is available upon request. Also, based on a 1 percent sub-sample in which optical character recognition software was used, a 99 percent correspondence was found with our measure of correctness based on the mode.

[14] The reason for this change of task is that the digitized data was part of a research project of another researcher who needed to digitize words from the census and not figures. The complexity relates to the quality of the pictures, which were more blurry, and therefore more difficult to type.

time, and a measure of *actual* working time (D_{it}/T_{it}), i.e. the fraction of contracted time that the worker actually worked.

The actual working time, D_{it}, differs from the *contractual* working time, T_{it}, in that workers do not work all the time. They can take *breaks* while remaining in the lab (i.e. coffee break, going to the bathroom, or just a period in which they do not type). Alternatively, they can leave the lab, which we label *absenteeism*. The actual working time is determined by discarding these periods of interruption (breaks plus absenteeism) from the contractual time. We cannot directly observe these interruptions in working time, but the software registered each moment when a character was typed. In the benchmark analysis we defined a break as an interruption period of more than 15 seconds between two subsequent instants when characters are typed.

Work interruptions can be further decomposed into breaks and absenteeism. We consider an additive decomposition in which the sum of fractions relative to contractual time (T_{it}) of actual working time (D_{it}), breaks (B_{it}), and periods of absenteeism (A_{it}) is equal to 1:

$$D_{it}/T_{it} + B_{it}/T_{it} + A_{it}/T_{it} = 1 \tag{4}$$

Ex ante **productivity**

To take out the selection effect from the total effect of productivity on-the-job, we use the answers from the productivity test the applicants took during the application process to construct the following average *ex ante* productivity measure:[15]

$$AP_{i,0}^{A} = C_i^A / D_i^A \tag{5}$$

[15] We only rely on the first test because it exactly corresponds to the task performed by workers under the period of study, and allows us to match as closely as possible the measures of *ex ante* and *ex post* productivity.

where the superscript A denotes that the figures refer to the *ex ante* productivity test, C_i^A is the number of correct characters that applicant i typed in test 1 of the productivity test, and D_i^A is the actual time that elapsed between typing the first and last characters in the test. We denote this time by D rather than T, because it is conceptually closer to D_{it} in the aforementioned decomposition formula of *ex post* productivity.

We also decompose the *ex ante* measure. However, since we cannot measure the contractual time T *ex ante*, we break it down into two terms instead of three:

$$AP_{i,0}^A \equiv C_i^A / D_i^A = \left(C_i^A / N_i^A\right) \times \left(N_i^A / D_i^A\right) \equiv AP_{i,1}^A \times AP_{i,2}^A \tag{6}$$

where N_i^A is the total number of characters typed in the productivity test. This decomposes the total *ex ante* productivity into a measure of *precision* and *speed*.

Descriptive statistics for applicants and hired workers

In this section, we provide descriptive statistics across treatment groups for the available explanatory variables for both the sample of applicants who took the productivity test and the hired workers who completed the experiment. We also check the extent to which the treatment groups are balanced within these samples. As the applicants are randomly assigned to the different treatments, this is a validity check of the randomization. Yet since each contract type will attract different types of workers, and as the eventual hiring of workers is the outcome of this self-selection process, we are interested in analyzing to what extent this balance disappears within the sample of *hired* workers, especially along the *ex ante* productivity measure.

We use the standardized difference (SD) to evaluate whether the samples are balanced across treatment groups (Rosenbaum and Rubin, 1985). The advantage of the SD is that it is not sensitive to sample size. This is particularly convenient in our context, because the sample of hired workers is so small that conventional t-tests are not appropriate. Following Rosenbaum and Rubin, we consider an SD exceeding 0.20 to be large.

The sample of applicants

Table 3 describes the sample of applicants. The first four columns report the means (and the associated SDs in parentheses) of the considered variables for each of the treatment groups. The last three columns report the SDs for treatments T2–T4 relative to T1.

The profession of data clerk is a routine job that typically requires a medium level of education, as some familiarity with information and communication technologies is required. We indeed find that nearly half of the applicants have a vocational degree, which in Colombia is a technical education. Students who complete this level of education receive a diploma that is valued more than a high-school degree, but less than a university degree. The temporary nature of the job explains why nearly half of the applicants are younger than 26 and about two-thirds are women. We also observe that 20 percent declared having dependents in the household, i.e. persons in need of care, and the same proportion were high-productivity types as measured by the pre-tests. The last three rows of the Table 3 present the averages of the *ex-ante* productivity measures presented above – i.e. efficiency, speed, and precision.

As expected, these variables are balanced across treatment groups.[16] Importantly, we see the *ex ante* measures of productivity are balanced across treatments. We only find

[16]It is noteworthy that the *number* of individuals assigned to the different treatments is not completely balanced. This is because we realized only after the random assignment that some individuals we retained only completed part of the productivity test; we dropped these individuals from the analysis. As demonstrated in the table, this did not result in any major selectivity in the sample.

an SD that exceeds 0.20 for the age group 26–30 when comparing T3 (full-time flexible) with the reference treatment T1: the share of applicants in T3 is 9 percentage points higher than the corresponding share in T1 (25 percent versus 16 percent). We will see in the robustness-checks section that our findings are robust to adding those as control variables.

The sample of hired workers

Table 4 describes the sample of hired workers. As expected, the variables are now imbalanced between the treatment groups, as the majority of the SDs are larger than 0.20.

We expected *a priori* that flexible work arrangements would especially attract women and individuals with caring responsibilities. We indeed observe that the most flexible contract (T4) attracted the highest share of individuals with dependents: 0.43 is more than twice as high as the average share in the sample. However, this flexible arrangement does not contain the highest share of women hired. The share of women is highest in the part-time non-flexible arrangement (T2), which contains a below-average share with dependents. The full-time flexible arrangement (T3) seems to attract relatively more women and individuals without caring responsibilities than the other types of contracts. We do not have a clear explanation for these deviations from our expectations. It might be related to the small sample size. We do not describe the selection patterns across the treatments for the other explanatory variables, because we do not have particular *a priori* hypotheses regarding their direction.

Finally, in the last three rows of the table, we observe significant differences in the *ex-ante* productivity measures across all treatment groups. Specifically, hired workers in treatment T1 (full-time non-flexible) are less efficient, slower, and more imprecise than workers in treatments T3 (full-time flexible) and T4 (part-time flexible) during the pre-employment test. This means that full-time flexible workers typed more numbers per minute during the test and the numbers were correctly typed more often. Hired workers under contract environment T2 (part-time non-flexible) are faster but less precise than under T1. These differences may induce bias in estimating the total effect of flexibility on *ex-post* productivity, since more productive workers are more likely to accept more flexible jobs, especially full-time flexible schedules. We will consider this imbalance in the estimation of the overall effect on productivity in the main results section, which in turn allows us to disentangle the selection from the motivational effect.

Ex post on-the-job productivity and working-time flexibility

Empirical strategy

In this section we analyze how different flexible working arrangements affect on-the-job observed productivity. We recorded information about the same data entry task during two contracted weeks using the different measures of productivity described above. First, we take into account the decomposition of productivity presented in equation (3), which becomes additive after the logarithm is taken, and then the decomposition proposed in equation (4) about work interruptions. In each of those analyses, we consider both the cases with and without controlling for measures of *ex ante* productivity. When we perform the analysis without controlling for *ex ante* productivity, we measure the overall impact of contract flexibility on the chosen *ex post* productivity measures, combining the selection effect and the behavioral effects. By controlling for the *ex ante* productivity measures, we seek to isolate the behavioral or motivational impact of contract flexibility, thus getting rid of the selection effect. We jointly control for two measures of *ex ante* productivity: the logarithm of the overall *ex ante* productivity measure ($AP^A_{i,0} \equiv C^A_i / D^A_i$) and the logarithm of the component measuring precision ($AP^A_{i,1} \equiv C^A_i / N^A_i$) as defined in equation (6).[17]

The benchmark regression without controlling for *ex ante* productivity is written as:

$$ln(Y_{it}) = \beta_1 + \beta_2 T_2 + \beta_3 T_3 + \beta_4 T_4 + \delta S_i + \lambda_2 I_{t2} + \epsilon_{it} \tag{7}$$

for $t \in \{1, 2\}$ where Y_{it} denotes one of the considered measures of observed productivity for individual i in week t, $I_{t2} = 1$ if $t = 2$ and $I_{t2} = 0$ if $t = 1$, so that λ_2 accounts for the differential productivity in week 2 relative to week 1, the reference. T_j for $j \in \{1, ..., 4\}$ and S_i are, respectively, indicators of the contract type and sampling strata, and ϵ_{it} denotes the zero-mean error of the regression.

When controlling for the *ex ante* productivity measures – overall productivity $ln(AP^A_{i,0})$ and precision $ln(AP^A_{i,1})$ – equation (7) becomes:

$$ln(Y_{it}) = \beta_1 + \beta_2 T_2 + \beta_3 T_3 + \beta_4 T_4 + \delta S_i + \lambda_2 I_{t2} + \gamma_1 ln(AP^A_{i,0}) + \gamma_2 ln(AP^A_{i,1}) + \epsilon_{it} \tag{8}$$

Estimating regression equations (7) and (8) entails two important challenges with respect to the inference of the parameters of interest. First, given the small sample size, we cannot rely on the standard asymptotic inference to test our hypotheses. We therefore use Fisher's exact inference to obtain the p-values of the considered null hypotheses (Imbens and Rubin, 2015). A second potential problem of inference is that we must

[17] The logarithm of the measure of speed is, according to equation (6), linearly dependent on the two other measures, and is therefore redundant as an additional control variable. The results are unchanged if we use a different pair of variables among the three.

consider the fact that productivity is correlated between the two weekly observations for each individual in the sample. A standard solution to this problem is to cluster the standard errors by individual. However, if there are only a small number of clusters, as is the case here, the standard errors are at risk of being biased downwards (Cameron and Miller, 2015). By relying on Fisher's exact inference, we can resolve this issue as well (Heß, 2017). Below we report the standard asymptotic standard errors for all coefficients, and the exact p-values for only the coefficients of interest. The reported significance of coefficients will be based on the latter, if available.

Main results

Table 5 reports the estimated effects of the different forms of contract flexibility in the benchmark model, equation (7). Column (1) corresponds to the baseline regression including only the stratification variables and the week fixed effects. In column 2, we augment the baseline regression with the measures from the *ex ante* productivity test conducted in the selection phase corresponding to the estimation of equation (8).

Considering the estimated parameters of interest in column (1), we find that only the full-time flexible contract has a significant effect (exact p-value of 5 percent) on the benchmark productivity measure. Workers hired with this contract are nearly 50 percent ($\exp(.40) = 1.49$) more productive than full-time non-flexible workers. For part-time workers under both flexible and non-flexible schedules, the effects on productivity are both positive, but these effects are not statistically different from that of full-time non-flexible workers.

The finding that a full-time flex contract attracts more productive workers than the other type of contract might be partly related to contextual factors. In Bogotá, more than half of the workforce is employed in precarious informal jobs with wages below the minimum wage. In such a context, workers prefer flexibility, as this is the norm to which they are accustomed. Moreover, workers may prefer a full-time to a part-time job because it pays more. This context may be different in a richer country.

The effects encompass the total and the selection effect of contract schedule flexibility. When controlling for selection effects in column (2), we find that there remains a role for an on-the-job motivational effect. According to the point estimate, productivity still increases relative to a full-time non-flexible contract by 32 percent ($\exp(.28) = 1.32$). This effect is significant at the 10 percent level. This estimate suggests that the on-the-job motivational effect represents an increase of about 60 percent of the total effect on productivity and the selection effect an increase of about 40 percent for full-time flexible jobs relative to full-time non-flexible contracts.

The effects of a part-time flexible and non-flexible contract are both positive, but much smaller and not significantly different from zero. We therefore cannot conclude that there is evidence that the flexibility of part-time work enhances productivity, even combined with a flexible working schedule.

Exploring mechanisms

We explore the mechanisms of the working arrangements on productivity by breaking down the benchmark productivity measure into different components. First, we consider the logarithmic additive form of the decomposition displayed by equation (3). Next, we consider the additive decomposition displayed in equation (4).

Decomposing the effects on global productivity

Columns (1) to (4) of Table 6 present the estimates of the first decomposition of the *global* effects on productivity, which comprise both the selection effects generated by the application process and the motivational on-the-job effects. In columns (5) to (8) the same decomposition is displayed after controlling for the *ex ante* productivity or selection effects. Columns (1) and (5) – denoted by "Total" – restate the total effect on productivity reported in Table 5. The subsequent columns present the coefficients of the decomposition.

We first focus on the findings for the workers on a full-time flexible contract, as we only find a significantly higher productivity for this contract type relative to full-time non-flexible contracts. The breakdown suggests that the motivational component, which represents about 60 percent of the total effect on productivity, is completely explained by more effective time use in the workplace. This follows from the fact that when selection effects are controlled for, the estimated coefficient explaining the effect on effective time use (0.25) is nearly as large as the one explaining the total productivity effect (0.28) (see columns (4) and (8)), while the coefficients that measure the reaction due to *precision* and *speed* are close to zero and not statistically significant. On the other hand, the selection process seemed to have attracted workers who have the ability to type faster, as the coefficients associated with these components are larger for speed ($0.14 > 0.04$). However, these are statistically insignificant.

Another interesting result is that workers on part-time non-flexible schedules also seem to demonstrate a more effective use of time than those in the comparison group (columns (4) and (8)), despite no observed differences in total productivity (columns (1) and (5)), and potentially at the loss of precision, though this loss is not significant (columns (2) and (6)). These effects should be interpreted as behavioral effects rather than selection effects, as the corresponding coefficients are of similar magnitude regardless of whether selection effects are controlled for or not.

The results from these specifications therefore provide evidence that working in more flexible environments makes individuals work more effectively, reducing their shirking time at work.

Our finding that the motivational effects of working-time schedules are largely driven by an increased number of working hours is in line with the findings of Beckmann et al. (2017), who use the German Socio-Economic Panel. However, our results seem to contradict those of Dutcher (2012), who reports that *workplace* flexibility has negative effects on productivity for a similar "dull" data entry task as the one in our study. A potential explanation is that workplace flexibility implies not only working-time flexibility, but also that the work is not monitored. For routine tasks, monitoring may be an essential device to keep workers committed to their work, as intrinsic motivation may be lacking (Beckmann, 2016). Yet workers may still want to reciprocate the provided working-time autonomy by increasing their effort (ibid). Gibbs et al. (2021) find an increase in the number of hours worked when high-skilled workers changed the workplace from the office to home. But the increase in working hours did not translate into more output per worker and it remained at the pre-COVID-19 level. The study shows that the extra hours worked were devoted to more coordination and meetings online. However, their results are valid for high-skilled tasks that require coordination.

Our findings prove that part-time workers on non-flexible contracts also reciprocate by effectively managing their working time, but this is partly counterbalanced by a lack of precision (though not significant). The latter is consistent with both the hypothesis that part-time workers are less effective than their full-time counterparts because entering data typically involves making more mistakes at the start (i.e. higher fixed start-up costs), and that part-time contracts attract less committed workers, a selection effect for which we cannot control using our *ex ante* measure of productivity (Künn-Nelen et al. (2013); Garnero (2016)). However, we do not find similar or stronger positive effects on effective working time for workers on *flexible* part-time contracts. The decomposition of the effect on working time discussed in the next subsection provides more insights into this result.

Decomposing the effects on time use

Table 7 presents the mean differences of the various uses of time across treatments for the global measure of productivity (columns (1) to (3)) and for *ex post* productivity after controlling for the selection effect (columns (4) to (6)). We estimate each component from equation 7 in order to evaluate which mechanism makes workers more productive at work, as a percentage of total contracted time. The additive decomposition allows us to disentangle the global effect of productivity on effective time D/T (worked time as a percentage of contracted time), absenteeism as a percentage of contracted time A/T, and time taken for breaks as a percentage of contracted time B/T.

We observe that selection effects play no role, as the magnitude of the effects on *ex post* productivity are not affected when we control for *ex ante* productivity. This is in line with our expectations, because working time refers to a motivational rather than a selection effect. The results also show that the positive productivity effects reported in the previous section are induced by the fact that workers with flexible work contracts take fewer breaks rather than being less absent.[18] [19]

A new result is that part-time workers with flexible contracts (T4) are found to reduce their break times by as much as 14 percentage points of the contractual working time, a reduction that is similar to that of part-time workers on non-flexible contracts (T2), and nearly twice as large as the effect for full-time workers on flexible contracts (T3). These findings suggest that part-time contracts *do* have motivational effects, but for various reasons these do not translate into higher global productivity. For part-time workers

[18]Recall that the effects are now measured as percentage-point effects relative to the contractual time, whereas in the previous tables proportional effects were measured on a logarithmic scale.

[19]A similar result was obtained by Bloom et al. (2015), who find that the majority of the increase in productivity is due to more hours worked.

on non-flexible contracts, this positive effect was tempered by the lower precision with which they enter data. For those on flexible contracts, this positive effect is largely absorbed by higher rates of absenteeism. A potential explanation is that workers with more caring responsibilities are much more likely to accept these kinds of contracts than others with more constraints. As the timing of care is very uncertain, they may be forced to be absent more than other workers. Another possibility is that they may be building their income by putting together many different types of jobs, so the opportunity cost of being absent is lower for them.

We also explored differences in start and finish times, noting that flexible workers are more likely to start work later than those on non-flexible work arrangements, the average difference amounting to 2 hours later for full-time workers, and nearly 3 hours for part-time workers, but that both groups also extend their finishing time by the same amount (see Table 8). This result is consistent with our suggestion that it is considerations to do with the use of time which will be the main explanatory mechanism behind our results.

Robustness checks

We estimate a robust-to-outliers regression, which is a weighted least squares regression that weighs the observations differently based on absolute residuals (Verardi and Croux, 2009). Column (2) of Table 9 displays a somewhat smaller effect of full-time flexible contracts on total productivity ($\exp(0.34) = 1.40$), but the effect remains statistically significant at the 5 percent level (Fisher exact p-value). When controlling for selection effects (column (4)), the coefficient is slightly reduced and the p-value increases slightly from 10 percent to 12 percent.

Results concerning the decomposition of time use are also robust and remain unchanged when using different measures of a break. In the benchmark analysis we defined a break as any period of interruption that lasted for more than 15 seconds between two subsequent instants when characters are typed. We tested whether results remained unchanged using alternative periods of interruption of more than 10, 20, 25, or 30 seconds. The results are reported in Table 10.

Conclusion

We conducted a field experiment to test whether working-time flexibility affects the productivity of workers hired for a temporary routine job. A novelty of our study is that we randomized job applicants (rather than already employed people) into different contractual environments and tested the productivity of these job seekers at the start of the application process before assigning them into these different environments. This allowed us to distinguish the selection effect of working-time flexibility from the total effect.

We found that both part-time work and working-time flexibility substantially reduced working time interruptions. However, for part-time workers this did not result in a statistically significant globally higher level of productivity, either because these workers also made more mistakes, or because they were more frequently absent from work. By contrast, for full-time workers who were given the autonomy to choose when to start and stop working, this enhanced effective working time was not counterbalanced by less precision or more absenteeism. Their productivity increased by as much as 30 percent. Furthermore, the working-time autonomy not only increase productivity thanks to these motivational effects, but also enhanced the attractiveness of the job for intrinsically more productive workers. This selection effect further increased productivity by 20%, so that eventual productivity was estimated to be as much as 50 percent higher than that of workers hired under a non-flexible full-time contract.

We have thus found that working-time flexibility can enhance productivity in temporary routine jobs, and we have decomposed our results according to selection and on-the-job motivational effects for new hires: these findings are new and may provide valuable advice for human resource strategies. This has strong policy implications: we show that introducing more working-time flexibility could be Pareto improving for both employers and employees. This is even more important because lower-wage workers often do not have access to this type of flexibility and are more likely to benefit from it (Kim, 2020). This is often a topic for debate for firms and for policy makers when thinking about how to offer better jobs to the low-skilled population and how this can be transformed into more productivity.

Moreover, we believe that our findings may generalize to other contexts, because the routine cognitive tasks that had to be executed in this experiment are very similar to the tasks required in several clerical support occupations. Indeed, Autor et al. (2003) define a composite measure of routine cognitive tasks, and find that occupations in clerical support, such as secretaries, cashiers, costumer services representatives, and filing clerks, are particularly heavy with regards to routine cognitive tasks. Mihaylov and Tijdens (2019) also identified occupations with routine tasks similar to those in this experiment, for which our results could be extrapolated. These are: record keeping, formatting correspondence, ordering stores, maintaining databases, dealing with incoming calls and messages, filing documents, compiling inventories, counting, and recording money. Nevertheless, our findings cannot be generalized to all types of jobs and institutional environments. In particular, the experiment was implemented in the Colombian labor market, where a very large share of the workforce is employed in informal jobs. Future research should scale up the experiment and extend it to other routine and non-routine tasks and working environments. Also, the experimental setting proposed in this study could be translated to a real-life environment, with a larger number of workers being hired under different work arrangements by a real firm.

We identified selection effects based on a measure of intrinsic productivity. However, we have seen that selection effects could be related to other intrinsic characteristics of applicants, such as commitment or caring responsibilities for dependents, which in turn may generate on-the-job motivational effects. These examples reveal that selection and behavioral effects are often very much intertwined. Disentangling these effects may therefore often be a matter of interpretation. While this suggests promising avenues of research, we believe the decompositions of the kind that we proposed here improve our understanding of why certain flexible working arrangements are effective (or not) in enhancing workers' productivity.

References

Allen, Tammy D, Ryan C Johnson, Kaitlin M Kiburz, and Kristen M Shockley, "Work–family conflict and flexible work arrangements: Deconstructing flexibility," *Personnel psychology*, 2013, *66* (2), 345–376.

Angelici, Marta and Paola Profeta, "Smart-Working: Work Flexibility without Constraints," *CESifo Working Paper Series*, 2020, (8165).

Autor, David H, Frank Levy, and Richard J Murnane, "The skill content of recent technological change: An empirical exploration," *The Quarterly journal of economics*, 2003, *118* (4), 1279–1333.

Beckmann, Michael, "Working-time autonomy as a management practice," *IZA World of Labor*, 2016, (206).

_ **, Thomas Cornelissen, and Matthias Kräkel**, "Self-managed working time and employee effort: Theory and evidence," *Journal of Economic Behavior & Organization*, 2017, *133*, 285 – 302.

Bloom, Nicholas, James Liang, John Roberts, and Zhichun Jenny Ying, "Does working from home work? Evidence from a Chinese experiment," *The Quarterly Journal of Economics*, 2015, *130* (1), 165–218.

Brescoll, Victoria L, Jennifer Glass, and Alexandra Sedlovskaya, "Ask and ye shall receive? The dynamics of employer-provided flexible work options and the need for public policy," *Journal of Social Issues*, 2013, *69* (2), 367–388.

Bustelo, Monserrat, Ana María Díaz Escobar, Jeanne Lafortune, Claudia Piras, Luz Magdalena Salas, José Tessada et al., "What is The Price of Freedom?: Estimating Women's Willingness to Pay for Job Schedule Flexibility," *Economic Development and Cultural Change*, 2022.

Cameron, Colin and Douglas Miller, "A Practitioner's Guide to Cluster-Robust Inference," *Journal of Human Resources*, 2015, *50* (2), 317–372.

Chen, Kuan-Ming, Claire Ding, John A List, and Magne Mogstad, "Reservation Wages and Workers' Valuation of Job Flexibility: Evidence from a Natural Field Experiment," Technical Report, National Bureau of Economic Research 2020.

Chen, M Keith, Peter E Rossi, Judith A Chevalier, and Emily Oehlsen, "The value of flexible work: Evidence from uber drivers," *Journal of Political Economy*, 2019, *127* (6), 2735–2794.

Chung, Heejung, "'Women's work penalty'in access to flexible working arrangements across Europe," *European Journal of Industrial Relations*, 2019, *25* (1), 23–40.

— **and Mariska Van der Horst**, "Flexible working and unpaid overtime in the UK: The role of gender, parental and occupational status," *Social Indicators Research*, 2018, pp. 1–26.

— **and** — , "Flexible working and unpaid overtime in the UK: The role of gender, parental and occupational status," *Social Indicators Research*, 2020, *151* (2), 495–520.

Dutcher, E Glenn, "The effects of telecommuting on productivity: An experimental examination. The role of dull and creative tasks," *Journal of Economic Behavior & Organization*, 2012, *84* (1), 355–363.

Eaton, Susan C, "If you can use them: Flexibility policies, organizational commitment, and perceived performance," *Industrial Relations: A Journal of Economy and Society*, 2003, *42* (2), 145–167.

Eurofound, "Sixth European Working Conditions Survey. Luxembourg: Publications Office of the European Union," 2015.

Garnero, Andrea, "Are part-time workers less productive and underpaid?," *IZA World of Labor*, 2016, (249).

Gibbs, Michael, Friederike Mengel, and Christoph Siemroth, "Work from Home & Productivity: Evidence from Personnel & Analytics Data on IT Professionals," *University of Chicago, Becker Friedman Institute for Economics Working Paper*, 2021, (2021-56).

Golden, Lonnie, "Flexible work schedules: Which workers get them?," *American Behavioral Scientist*, 2001, *44* (7), 1157–1178.

—, "Irregular work scheduling and its consequences," *Economic Policy Institute Briefing Paper*, 2015, (394).

—, Stephen Sweet, and Heejung Chung, "Positive and negative application of flexible working time arrangements: comparing the United States of America and the EU countries," in "Handbook of Research on Comparative Human Resource Management," Edward Elgar Publishing, 2018.

Harrington, Emma and Natalia Emanuel, "'Working' Remotely? Selection, Treatment, and Market Provision of Remote Work (JMP)," *Harvard University,Job Market Paper*, 2020.

He, Haoran, David Neumark, and Qian Weng, "Do Workers Value Flexible Jobs? A Field Experiment on Compensating Differentials," *Journal of Labor Economics - forthcoming*, 2021.

Henly, Julia R and Susan J Lambert, "Unpredictable work timing in retail jobs: Implications for employee work–life conflict," *Ilr Review*, 2014, *67* (3), 986–1016.

Heß, Simon, "Randomization Inference with Stata: A Guide and Software," *The Stata Journal*, 2017, *17* (3), 630–651.

Imbens, Guido W. and Donald B. Rubin, *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*, Cambridge University Press, 2015.

Kamalahmadi, Masoud, Qiuping Yu, and Yong-Pin Zhou, "Call to duty: Just-in-time scheduling in a restaurant chain," *Management Science*, 2021.

Kelliher, Clare and Deirdre Anderson, "Doing more with less? Flexible working practices and the intensification of work," *Human relations*, 2010, *63* (1), 83–106.

_ **and Lilian M De Menezes**, *Flexible working in organisations: A research overview*, Routledge, 2019.

Kim, Jaeseung, "Workplace flexibility and parent–child interactions among working parents in the US," *Social Indicators Research*, 2020, *151* (2), 427–469.

Kossek, Ellen and Jesse Michel, "Flexible work schedules," *APA Handbook of Industrial and Organizational Psychology*, 01 2011, *1*, 535–572.

Künn-Nelen, Annemarie, Andries De Grip, and Didier Fouarge, "Is part-time employment beneficial for firm productivity?," *ILR Review*, 2013, *66* (5), 1172–1191.

Lambert, Susan J, "Passing the buck: Labor flexibility practices that transfer risk onto hourly workers," *Human relations*, 2008, *61* (9), 1203–1227.

Leslie, Lisa M, Colleen Flaherty Manchester, Tae-Youn Park, and Si Ahn Mehng, "Flexible work practices: a source of career premiums or penalties?," *Academy of Management Journal*, 2012, *55* (6), 1407–1428.

Lott, Yvonne and Heejung Chung, "Gender discrepancies in the outcomes of schedule control on overtime hours and income in Germany," *European Sociological Review*, 2016, *32* (6), 752–765.

Magnusson, Charlotta, "Flexible time–but is the time owned? Family friendly and family unfriendly work arrangements, occupational gender composition and wages: a test of the mother-friendly job hypothesis in Sweden," *Community, Work & Family*, 2021, *24* (3), 291–314.

Mas, Alexandre and Amanda Pallais, "Valuing alternative work arrangements," *American Economic Review*, 2017, *107* (12), 3722–59.

Mihaylov, Emil and Kea Gartje Tijdens, "Measuring the routine and non-routine task content of 427 four-digit ISCO-08 occupations," 2019.

Nagin, Daniel S., James B. Rebitzer, Seth Sanders, and Lowell J. Taylor, "Monitoring, Motivation, and Management: The Determinants of Opportunistic Behavior in a Field Experiment," *American Economic Review*, September 2002, *92* (4), 850–873.

Nelen, Annemarie, Andries De Grip, and Didier Fouarge, "Is part-time employment beneficial for firm productivity?," *Available at SSRN 1741611*, 2011.

OECD, "Background brief on how workplace flexibility can help European employees to balance work and family," 2016.

Ortega, Jaime, "Why do employers give discretion? Family versus performance concerns," *Industrial Relations: A Journal of Economy and Society*, 2009, *48* (1), 1–26.

Pierce, Jon L and John W Newstrom, "The design of flexible work schedules and employee responses: Relationships and process," *Journal of Occupational Behaviour*, 1983, pp. 247–262.

Rosenbaum, Paul R and Donald B Rubin, "Constructing a control group using multivariate matched sampling methods that incorporate the propensity score," *The American Statistician*, 1985, *39* (1), 33–38.

Rupietta, Kira and Michael Beckmann, "Working from Home: What is the Effect on Employees' Effort?," *Schmalenbach Business Review*, 12 2017, *70.*

Schneider, Daniel and Kristen Harknett, "Consequences of Routine Work-Schedule Instability for Worker Health and Well-Being," *American Sociological Review*, 2019, *84* (1), 82–114. PMID: 33311716.

_ and _, "Hard Times: Routine Schedule Unpredictability and Material Hardship among Service Sector Workers," *Social Forces*, 2021, *99* (4), 1682–1709.

Spreitzer, Gretchen M, Lindsey Cameron, and Lyndon Garrett, "Alternative work arrangements: Two images of the new world of work," *Annual Review of Organizational Psychology and Organizational Behavior*, 2017, *4*, 473–499.

Verardi, Vincenzo and Christophe Croux, "Robust regression in Stata," *The Stata Journal*, 2009, *9* (3), 439–453.

Williams, Joan C, Saravanan Kesavan, and Lisa McCorkell, "When retail workers have stable schedules, sales and productivity go up," *Harvard Business Review*, 2018.

Wiß, Tobias, "Paths towards family-friendly working time arrangements: Comparing workplaces in different countries and industries," *Social policy & administration*, 2017, *51* (7), 1406–1430.

Tables and Figures

Table 1. Samples at Each Stage of the Experiment

Sample	*Experimental stages*	*N*
0	Applicants who started the online survey	686
1	*First randomization: assignment of contract types*	
	Applicants who finished the survey and tests, & received 1^{st} email with the contract type	535
2	Applicants who stated interest after 1^{st} email containing the contract type	438
3	Applicants who stated interest after 2^{nd} email asking for new open questions	384
4	*Second randomization: who gets an offer among each contract type*	
	Applicants who received an offer	79
	Applicants who received an offer and accepted it, and were hired	38
	Applicants who took the job and finished it	34

Table 2. Job Offers and Acceptance Rate

	Offers	Acceptance Rate (%)	Workers	(%)
T1: *Full-time non-flex*	17	70.6	11	32.35
T2: *Part-time non-flex*	19	42.1	7	20.59
T3: *Full-time flex*	23	40.5	9	26.47
T4: *Part-time flex*	20	40	7	20.59
Total	79	100	34	100

Figure 1. Task Sample

PLACE FIGURE 1 (MAIN) HERE

Table 3. Descriptive Statistics by Treatment Group: Sample of Applicants

Variable	(T1) Full-time non-flexible Mean (1)	(T2) Part-time non-flexible Mean (2)	(T3) Full-time flexible Mean (3)	(T4) Part-time flexible Mean (4)	Standardized differences (2)-(1)	(3)-(1)	(4)-(1)
Stratification Variables							
Female	0.69	0.63	0.67	0.65	-0.12	-0.04	-0.08
	(0.47)	(0.48)	(0.47)	(0.48)			
High Productivity	0.23	0.18	0.19	0.23	-0.12	-0.10	0.00
	(0.42)	(0.39)	(0.39)	(0.42)			
Dependents	0.24	0.20	0.19	0.19	-0.08	-0.12	-0.12
	(0.43)	(0.40)	(0.39)	(0.39)			
Additional Control Variables							
Age 20–25 yr old	0.47	0.44	0.43	0.50	-0.06	-0.09	0.07
	(0.50)	(0.50)	(0.50)	(0.50)			
Age 26–30 yr old	0.16	0.23	0.25	0.22	0.16	0.22	0.14
	(0.37)	(0.42)	(0.44)	(0.42)			
Age 31 and more	0.36	0.33	0.32	0.27	-0.06	-0.09	-0.19
	(0.48)	(0.47)	(0.47)	(0.45)			
Education: High School or less	0.26	0.29	0.32	0.24	0.05	0.12	-0.06
	(0.44)	(0.45)	(0.47)	(0.43)			
Education: University	0.31	0.29	0.23	0.27	-0.04	-0.18	-0.09
	(0.46)	(0.45)	(0.42)	(0.44)			
Education: Vocational	0.43	0.42	0.45	0.50	-0.01	0.05	0.13
	(0.50)	(0.50)	(0.50)	(0.50)			
Ex ante productivity measures							
Efficiency(C/T) Test 1	0.09	0.09	0.09	0.09	-0.05	-0.03	0.06
	(0.04)	(0.04)	(0.05)	(0.04)			
Speed (N/T) Test 1	0.11	0.11	0.12	0.12	0.02	0.08	0.12
	(0.04)	(0.05)	(0.05)	(0.04)			
Precision(C/N) Test 1	0.81	0.79	0.78	0.81	-0.13	-0.19	-0.02
	(0.19)	(0.21)	(0.19)	(0.17)			
N	140	132	150	113			

Notes: The first four columns present the mean value of the covariates in the corresponding treatment groups. The last three columns report the standardized differences defined as $(\bar{X}_j - \bar{X}_{T1})/\sqrt{\left[\left(Var(X_j) + Var(X_{T1})\right)/2\right]}$, where $\bar{X}_j$ and $Var(X_j)$ are the sample mean and the variance of X_j for $j \in \{T2, T3, T4\}$. Following Rosenbaum and Rubin (1985), we consider an SD exceeding 0.20 to be unbalanced.

Table 4. Descriptive Statistics by Treatment Group: Sample of Workers

Variable	(T1) Full-time non-flexible Mean (1)	(T2) Part-time non-flexible Mean (2)	(T3) Full-time flexible Mean (3)	(T4) Part-time flexible Mean (4)	Standardized difference (2)-(1)	(3)-(1)	(4)-(1)
Stratification Variables							
Female	0.64	0.86	0.44	0.57	0.48	-0.38	-0.13
	(0.50)	(0.38)	(0.53)	(0.53)			
High Productivity	0.27	0.14	0.22	0.14	-0.30	-0.11	-0.30
	(0.47)	(0.38)	(0.44)	(0.38)			
Dependents	0.27	0.14	0.11	0.43	-0.30	-0.39	0.32
	(0.47)	(0.38)	(0.33)	(0.53)			
Additional Control Variables							
Age 20–25 yr old	0.27	0.14	0.22	0.43	-0.30	-0.11	0.32
	(0.47)	(0.38)	(0.44)	(0.53)			
Age 26–30 yr old	0.36	0.43	0.11	0.43	0.13	-0.57	0.13
	(0.50)	(0.53)	(0.33)	(0.53)			
Age 31 and more	0.36	0.43	0.67	0.14	0.13	0.59	-0.48
	(0.50)	(0.53)	(0.50)	(0.38)			
Educ: High School	0.27	0.43	0.22	0.14	0.32	-0.11	-0.30
or less	(0.47)	(0.53)	(0.44)	(0.38)			
Education: Univer-	0.45	0.43	0.22	0.14	-0.05	-0.47	-0.64
sity	(0.52)	(0.53)	(0.44)	(0.38)			
Educ: Vocational	0.27	0.14	0.56	0.71	-0.30	0.56	0.86
	(0.47)	(0.38)	(0.53)	(0.49)			
Ex ante productivity measures							
Efficiency(C/T)	0.08	0.09	0.10	0.10	0.24	0.57	0.57
Test 1	(0.04)	(0.05)	(0.05)	(0.04)			
Speed (N/T) Test 1	0.10	0.11	0.12	0.14	0.33	0.58	0.68
	(0.02)	(0.04)	(0.06)	(0.09)			
Precision(C/N)	0.76	0.76	0.84	0.78	-0.02	0.40	0.08
Test 1	(0.23)	(0.31)	(0.16)	(0.12)			
N	11	7	9	7			

Notes: The first four columns present the mean value of the covariates in the corresponding treatment groups. The last three columns report the standardized differences defined as $(\bar{X}_j - \bar{X}_{T1})/\sqrt{\left[\left(Var(X_j) + Var(X_{T1})\right)/2\right]}$, where $\bar{X}_j$ and $Var(X_j)$ are the sample mean and the variance of X_j for $j \in \{T2, T3, T4\}$. Following Rosenbaum and Rubin (1985), we consider an SD exceeding 0.20 to be unbalanced.

Table 5. Treatment Effects on Ex-Post Productivity

Productivity (ln)	(1)	(2)
T2: Part-time non-flexible	.12	.18
	(.27)	(.18)
	[.66]	[.30]
T3: Full-time flexible	.40**	.28*
	(.15)	(.16)
	[.05]	[.08]
T4: Part-time flexible	.15	.09
	(.22)	(.23)
	[.18]	[.39]
Ref: T1 Full-time non-flexible		
Constant	–1.73***	–1.64***
	(.11)	(.49)
R Squared	.2	.52
N	68	68
Controls	No	Yes

Notes: Standard errors in parentheses are clustered at the treatment level. Exact p-value for the Fisher Tests in brackets defined as the proportion of possible treatment assignments that yield a t-statistic greater or equal to the observed t-statistics. 3000 permutations. Stratification dummies added to all regressions. Column (2) controls for the log of *ex-ante* productivity (*precision*), and log of *ex-ante total* productivity. Significance level: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. The full set of coefficients can be found in the supplemental material Table A2.

Table 6. Decomposition of the Total Effects on *Ex-Post* Productivity

	Global- Productivity				Without Selection Effects			
	Total	Precision	Speed	Effective Time	Total	Precision	Speed	Effective Time
	(C/T)	(C/N)	(N/D)	(D/T)	(C/T)	(C/N)	(N/D)	(D/T)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
T2: Part-time	.12	–.08	.00	.20	.18	–.07	–.00	.26*
non flexible	(.27)	(.10)	(.14)	(.12)	(.18)	(.05)	(.10)	(.13)
	[.66]	[.35]	[.98]	[.27]	[.30]	[.37]	[.98]	[.14]
T3: Full-time	.40**	.04	.14	.22**	.28**	–.01	.04	.25**
flexible	(.15)	(.02)	(.08)	(.12)	(.16)	(.04)	(.10)	(.11)
	[.05]	[.68]	[.15]	[.05]	[.08]	[.90]	[.58]	[.09]
T4: Part-time	.15	.03	.12**	.00	.09	–.01	.03	.07
flexible	(.22)	(.02)	(.11)	(.17)	(.23)	(.04)	(.13)	(.13)
	[.18]	[.40]	[.07]	[0.99]	[.439]	[.73]	[.33]	[.65]
Ref: T1 Full-time non-flexible								
Constant	–5.75*	–.27	–3.47***	–2.02	–5.71*	–.13	–2.94***	–2.64
	(2.90)	(.32)	(.82)	(2.72)	(3.02)	(.38)	(.94)	(2.65)
R Squared	.2	.17	.29	.22	.52	.64	.54	.35
N	68	68	68	68	68	68	68	68

Notes: Standard errors in parentheses are clustered at the treatment level. Exact p-value for the Fisher Tests in brackets defined as the proportion of possible treatment assignments that yield a t-statistic greater or equal to the observed t-statistics. 3000 permutations. Stratification dummies included in all columns. Columns 5 to 8 control for the log of *ex-ante* productivity (*precision*), and for the log of *ex-ante total* productivity. *** p<0.01, ** p<0.05, * p<0.1. The full set of coefficients can be found in the supplemental material Table A3.

Table 7. Decomposition of the Time Use Effects

	Global			Without Sorting Effects		
	Effective Time (D/T) (1)	Absenteeism (A/T) (2)	Breaks (B/T) (3)	Effective Time (D/T) (4)	Absenteeism (A/T) (5)	Breaks (B/T) (6)
T2: Part-time non-flexible	.10	.04	–.14***	.12*	.03	–.15***
	(.06)	(.04)	(.04)	(.06)	(.04)	(.04)
	[.22]	[.32]	[.02]	[.08]	[.55]	[.01]
T3: Full-time flexible	.09	–.01	–.08	.10*	–.01	–.08**
	(.05)	(.04)	(.03)	(.05)	(.03)	(.03)
	[.19]	[.85]	[.12]	[.06]	[.74]	[.05]
T4: Part-time flexible	.02	.11	–.14***	.05	.10	–.14***
	(.07)	(.05)	(.04)	(.06)	(.03)	(.04)
	[.65]	[.30]	[.03]	[.17]	[.45]	[.00]
Ref: T1 Full-time non-flexible						
Constant	.43***	.05*	.51***	.22	.20**	.58***
	(.03)	(.03)	(.02)	(.15)	(.10)	(.09)
R Squared	.23	.26	.46	.35	.32	.48
N	68	68	68	68	68	68

Notes: Standard errors in parentheses are clustered at the treatment level. Exact p-value for the Fisher Tests in brackets defined as the proportion of possible treatment assignments that yield a t-statistic greater or equal to the observed t-statistics. 3000 permutations. Stratification dummies are included in all columns. Columns (4) to (6) control for the log of *ex-ante* productivity. *** p<0.01, ** p<0.05, * p<0.1. The full set of coefficients can be found in the supplemental material Table A4.

Table 8. Mean differences in starting and and finishing time

	non-flexible	flexible	T-test
Panel A: Full time			
	1	0	Difference
Variable	Mean/SE	Mean/SE	(1)-(2)
Starting time	9.16	11.20	-2.04***
	(0.20)	(0.27)	
Finishing time	16.05	18.13	-2.08***
	(0.09)	(0.17)	
N	103	80	
Panel B: Part time			
Starting time	9.68	12.45	-2.77***
	(0.32)	(0.50)	
Finishing time	12.65	15.80	-3.15***
	(0.29)	(0.45)	
N	58	52	

Notes: The value displayed for t-tests are the differences in the means across the groups. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table 9. Robustness Checks

	Robustness			
	Global		Without Sorting	
Productivity (ln)	(1)	(2)	(3)	(4)
T_2 Part-time no flexible	.12	.29	.17	.18
	(.27)	(.15)	(.19)	(.15)
	[.66]	[.19]	[.37]	[.40]
T_3 Full-time flexible	.40**	.34**	.28*	.28
	(.15)	(.14)	(.16)	(.15)
	[.06]	[.05]	[.10]	[.12]
T_4 Part-time flexible	.15	.08	.09	.07
	(.22)	(.15)	(.23)	(.16)
	[.17]	[.57]	[.61]	[.70]
Ref: T_1 Full-time non-flexible				
Ex ante productivity - *Precision*			.89***	.90***
			(.14)	(.22)
Ex ante total productivity			-0.08	–.06
			(.21)	(.15)
Constant	–1.73***	–6.89	-1.64***	–6.50
	(.11)	(5.03)	(.49)	(4.85)
R Squared	.2	.2	.52	. .48
N	68	68	68	68

Notes: The outcome variable is the log of productivity, measured as the number of correct digits over the contracted time. Standard errors in parentheses are clustered at the individual level. Exact p-value for the Fisher Tests in brackets defined as the proportion of possible treatment assignments that yield a t-statistic greater or equal to the observed t-statistics. 3000 permutations. Stratification variables are included in all columns. *** p<0.01, ** p<0.05, * p<0.1.

Table 10. Robustness Checks: change in the definition of a break

	Breaks				
	$\geq$ 10 sec	$\geq$ 15 sec	$\geq$ 20 sec	$\geq$ 25 sec	$\geq$ 30 sec
T2: Part-time non-flexible	.18	.18	.18	.18	.18
	(.18)	(.18)	(.18)	(.18)	(.18)
	[.31]	[.29]	[.29]	[.30]	[.30]
T3: Full-time flex	.28*	.28*	.28*	.28*	.28*
	(.16)	(.16)	(.16)	(.16)	(.16)
	[.09]	[.09]	[.08]	[.08]	[.08]
T4: Part-time flex	.09	.09	.09	.09	.09
	(.24)	(.23)	(.23)	(.23)	(.23)
	[.38]	[.42]	[.39]	[.39]	[.41]
Ref: T_1 Full-time non-flexible					
ln(Precision(C/N) Test 1)	.92***	.89***	.89***	.89***	.89***
	(.26)	(.26)	(.26)	(.26)	(.26)
ln(Efficiency(C/T) Test 1)	–.08	–.08	–.08	–.08	–.08
	(.21)	(.21)	(.21)	(.21)	(.21)
Week fixed effects	.08	.08	.08	.08	.08
	(.06)	(.06)	(.06)	(.06)	(.06)
Men/No dep/High prod	–.12	–.11	–.11	–.11	–.11
	(.18)	(.17)	(.17)	(.17)	(.17)
Men/Dep/Low prod	.40*	.39*	.39*	.39*	.39*
	(.20)	(.19)	(.19)	(.19)	(.19)
Women/No dep/Low prod	.08	.07	.07	.07	.07
	(.20)	(.20)	(.20)	(.20)	(.20)
Women/No dep/High prod	–.04	–.04	–.04	–.04	–.04
	(.14)	(.14)	(.14)	(.14)	(.14)
Women/Dep/Low prod	.46*	.46*	.46*	.46*	.46*
	(.25)	(.25)	(.25)	(.25)	(.25)
Women/Dep/High prod	.07	.07	.07	.07	.07
	(.15)	(.15)	(.15)	(.15)	(.15)
Constant	–5.74*	–5.71*	–5.71*	–5.71*	–5.71*
	(3.14)	(3.03)	(3.02)	(3.02)	(3.02)
R Squared	.53	.52	.52	.52	.52
N	68	68	68	68	68

Notes: The outcome variable is the log of productivity, measured as the number of correct digits over the contracted time. Standard errors in parentheses are clustered at the individual level. Exact p-value for the Fisher Tests in brackets. This exact p-value is the proportion of possible treatment assignments that yield a t-statistic greater than or equal to the observed t-statistics. We used 3000 permutations. Significance level: *** p<0.01, ** p<0.05, * p<0.1.