

Performance measures and worker productivity

Choosing the right performance measures can inform and improve decision-making in policy and management

Keywords: worker productivity, performance measures, human resource management

ELEVATOR PITCH

Measuring workers' productivity is important for public policy and private-sector decision-making. Due to a lack of reliable methods to determine workers' productivity, firms often use specific performance measures, such as how different incentives affect employees' behavior. The public sector also uses these measures to monitor and evaluate personnel, such as teachers. To select the right performance measures, and as a result design better employment contracts and improve productivity, policymakers and managers need to understand the advantages and disadvantages of the available metrics.

KEY FINDINGS

Pros

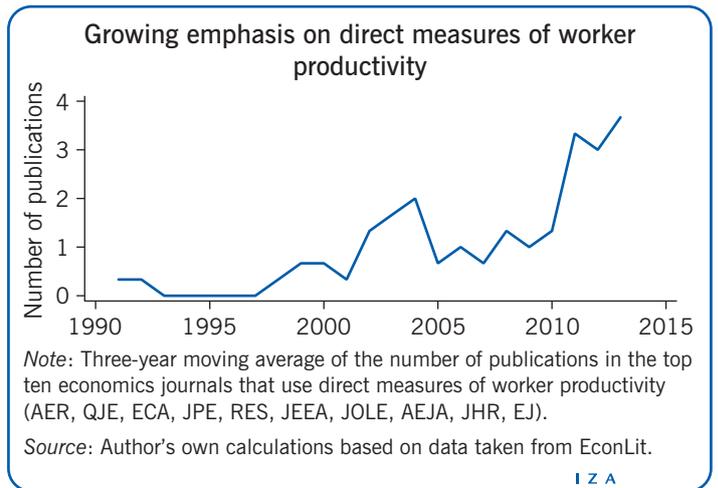
- ⊕ Performance measures provide detailed information about worker productivity.
- ⊕ To inform about a wide range of questions, such as how incentives work, how peer effects operate, or how workers accumulate human capital, performance measures can be useful.
- ⊕ Reliable performance measures are needed to design appropriate contracts and improve productivity.
- ⊕ Performance measures are increasingly available for low- and high-skilled jobs, as well as for jobs in the private and public sectors.

Cons

- ⊖ There is no universal definition of worker productivity; measures of worker productivity typically depend on the setting in which they are collected.
- ⊖ Worker productivity is usually multidimensional, but it is generally not possible to measure all dimensions.
- ⊖ If the wrong performance measures are chosen to evaluate workers, distortions can create negative effects on worker productivity.
- ⊖ For settings in which performance is only observable at the team level, it is not always possible to estimate individual contributions to team productivity.

AUTHOR'S MAIN MESSAGE

Measures of worker productivity can give important insights into how workers perform and how workplaces should be organized. Direct measures of productivity are used to study a range of questions, such as the effects of incentives on workers' productivity, the influence of peers on behavior, or the accumulation of human capital on the job. For these and related questions, it is important to select appropriate performance measures. This choice is critical, as relying on inappropriate measures can lead to the design of inefficient incentives, poor employment contracts, or wrong policy conclusions.



MOTIVATION

Politicians and managers often make decisions that involve the behavior of individuals at work. For example, managers decide how to establish optimal incentives or how much training to provide for employees, while policymakers make decisions about the regulation of working hours. To make informed choices about these kinds of issues, it is important to know how these decisions will affect workers' behavior, in terms of their productivity.

The conceptualization of worker productivity has gained increasing attention in the last decade. Direct measures are now commonly used in research within economics and related fields; they frequently serve as approximations of workers' productivity. This paper describes how worker productivity can be defined and provides an overview of the most up-to-date performance measurements available in order to help decision makers choose the right ones for their specific purposes.

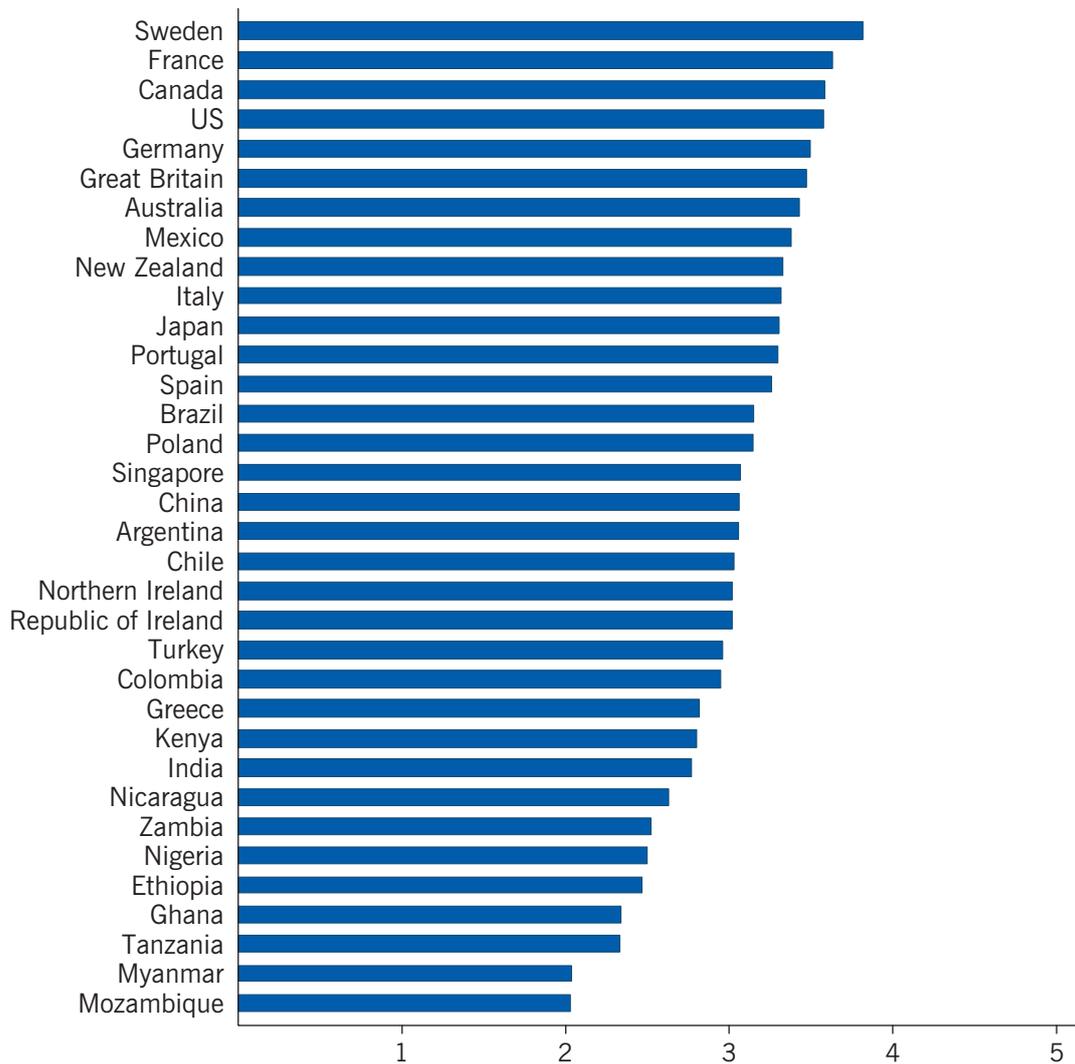
DISCUSSION OF PROS AND CONS

Defining worker productivity

In a general sense, productivity can be defined as the ratio between a measure of output and a measure of input. The productivity of workers could thus be measured as an output, e.g. sales or units produced, relative to an input, e.g. the number of hours worked or the cost of labor. Traditionally, labor productivity is derived from aggregate measures at the firm level, e.g. value-added per worker. To account for differences between labor inputs, this measure has often been disaggregated according to various labor types, e.g. low-, medium-, and high-skilled labor. However, even at this disaggregated level, measures of labor productivity can mask considerable variation with respect to workers' underlying productivity, either between workers, or over time. At the individual (worker) level, studies frequently use input measures, such as workers' wages, as a measure of productivity. Although correlated with the underlying productivity of the individual worker, there are several reasons why wages do not directly reflect the worker's actual productivity. For instance, institutional settings, such as those resulting from collective agreements, often make wages dependent on age or tenure rather than productivity. This is complicated by the fact that most data do not contain information on hourly wages, but rather on monthly wages. Variations in monthly wages might not only reflect differences in productivity, but also in the number of working hours. Furthermore, wage growth is often determined by supervisor evaluations, which might reflect bias due to gender or migration background.

Both labor productivity and wages have their shortcomings when it comes to assessing workers' productivity. Ideally, one would like to observe productivity for each individual worker at each point in time. In reality, however, output is rarely observable at the individual level for a reasonable cost, thus making it practically impossible to calculate each individual's productivity. Instead, firms use individual measures of workers' performance as an approximation of their productivity. Most occupations have one or more metrics that can be used to evaluate how well workers perform. These measures, also known as "key performance indicators" (KPIs), are regularly used for internal evaluation and monitoring in firms. Figure 1 shows evidence from the World Management Survey on the extent to which firms collect performance measures and use them for monitoring purposes.

Figure 1. Use of direct performance measurement across countries



Note: Firms' managers are asked to assess whether performance is tracked using meaningful metrics and with appropriate regularity, measured on a scale from 1 (lowest) to 5 (highest).

Source: World Management Survey, 2004–2014. Online at: <http://worldmanagementsurvey.org>

There is a key set of properties that should be met when assessing measures of worker productivity:

- (1) *Objectivity*: measures should be objective, as opposed to subjective, with respect to supervisor or peer ratings.
- (2) *Availability*: measures need to be available at the individual (worker) level (i.e. not on aggregated levels such as team or firm level).
- (3) *Comparability*: tasks and measurements should be the same across workers and time.
- (4) *Quality and controllability*: workers should have sufficient influence on the outcome, i.e. by choosing their own effort levels.

The precise measurement of performance across workers and over time allows policymakers to address important economic questions, such as how incentives affect workers' behavior,

how the presence of peers affects workers' productivity, or how workers accumulate human capital in firms.

Figure 2 provides examples of workers' productivity measures which were used in studies published in leading economics journals, and whose findings can be used to inform policymakers and practitioners. The figure shows that performance is not only measurable for low-skilled jobs with routine tasks, but that it can also be measured for rather knowledge-intensive, non-routine professions, such as lawyers, physicians, or scientists.

Figure 2. Examples of measures of workers' productivity

<i>Sector</i>	<i>Job/task</i>	<i>Frequency</i>	<i>Direct measure or estimated?</i>	<i>Quantity or quality dimension</i>
Service sector	Windshield installers: units installed per day [1], [2]	Days	Firm's database	Quantity
	Supermarket cashiers: number of items scanned per second [3]	Seconds	Firm's database	Quantity
	Call agents: average length of calls [4]	Weeks	Firm's database	Quantity (quality measure available)
	Lawyers: number of hours billed, new client revenue [5]	Years	Survey-based	Quantity and quality
Agriculture	Fruit pickers: kilograms of fruits picked per day [6]	Days	Firm's database	Quantity
Education and research	Academic scientists: PhD thesis gets published in top journal, number of citations [7]	–	University archives	Quantity and quality
	Academic scientists: impact factor weighted publications [8]	–	Public databases	Quality adjusted output
	Teachers: teacher value-added [9], [10]	Annual	Estimated	Quality adjusted output
Health care sector	Paramedics: transport time, number of procedures [11]	Shifts (daily)	Paramedics' firm database	Quantity and quality
Politics and public sector	Politicians: number of submitted bills [12]	–	Governmental survey	Quantity and quality
	Navy recruiters: number and quality of recruits [13]	Monthly	Navy data	Quantity and quality

Source: For full source details, see the complete list of references for this article at: <http://wol.iza.org/articles/performance-measures-and-worker-productivity>

Advantages and disadvantages of measuring workers' productivity

Firms regularly use measures of workers' performance to approximate productivity. Measuring workers' performance, however, is not always straightforward, or even possible, at a reasonable cost. Even though most occupations include some measures of performance

at the worker level, Figure 2 shows that there is no universal measure. Instead, the degree to which performance can be monitored depends on the setting.

What do performance measures entail?

Worker performance can be a function of many features, including the worker's effort, education, age, or tenure, and the firm's characteristics, such as work environment, wages, or incentives. An observed change in a worker's performance might be due to several reasons, including factors outside of the worker's control. The two most common reasons are changes in the worker's skills, e.g. due to training programs or from learning on-the-job, and changes in effort provided by the worker, e.g. due to different incentives set by the management. Technological change would also be a relevant subject to examine, but due to the typically short time horizons taken in the available studies (usually a matter of weeks), technology is generally considered as a constant in these cases.

A performance measure's usefulness for assessing workers' behavior crucially depends on the degree to which the worker has influence on the measure. Measures will be unreliable predictors of workers' productivity if they are largely driven by factors that are outside of a worker's control, such as variation in customer demand, or weather conditions in agriculture. Although any performance measure will contain some random variation, appropriate measures of workers' productivity should be balanced with respect to the determinants that are within the worker's control versus those that are not.

Multidimensional performance

In reality, it is difficult to assess workers' productivity using just one measure. Workers' jobs can include one or several tasks. University professors conduct research, are involved in teaching, and perform administrative tasks. Each of these, in turn, can be evaluated along different dimensions, e.g. by the quality and quantity of a task (workers could work quickly, but provide low quality, or slowly, but with high quality). Workers could be evaluated with separate performance measures for each relevant dimension. The task of conducting research, for example, could be measured by the number of publications, but also by the quality of the publications, e.g. measured by a journal's impact factor. Although quality and quantity are dimensions that apply to almost every task, one could also think of other dimensions, e.g. the policy relevance of the research.

Even for single-task occupations, the multi-dimensionality of workers' productivity has important implications with respect to the use of performance measures for policy determination. Whereas the firm is interested in productivity, workers are usually evaluated (and incentivized) based on specific performance measures. Since workers' productivity is not perfectly observable, and given that human resource decisions are frequently based on observable performance measures, distortions can arise. This would be the case, for example, if a firm only observes how much total output a worker produces, but ignores the quality of the output. Incentives based on observable measures could result in distortion, i.e. that the incentives of the firm and the worker are not perfectly aligned. Whereas the firm is interested in both dimensions, the worker might solely focus on performing well on the measurable performance indicators, which does not necessarily coincide with the un-measurable ones. While it may be possible to mitigate this problem through intrinsic motivation, distortions are likely to create negative effects if incentive schemes are not well designed.

One way to get around this incentive problem is to use alternative measures, such as firm-level valued-added metrics. While these types of measures appear to solve the problem of distortion, additional risks arise for the worker, since these measures are less controllable from the individual's point of view. This issue demonstrates an important trade-off between distortion on the one hand, and how precisely a performance measure can actually assess workers' effort, on the other.

Another alternative to one-dimensional measures is subjective performance measures, such as supervisor ratings. Although these types of evaluation are often only observable on an annual basis, and might contain subjectivity bias, they can capture a broader picture of performance than single measures allow.

Quantity and quality

It is difficult to identify appropriate measures for multiple performance dimensions. Two very common dimensions are the quality at which a job is performed, and how fast it is done (quantity). If, for example, workers' productivity in manufacturing is analyzed by means of a quantity-related measure of output, such as the number of products manufactured per hour, then a suitable candidate to measure the quality of output could be the number of defective products produced, i.e. the defection rate.

Figure 2 shows a range of studies that primarily use performance measures showing how quickly a worker performs (i.e. quantity-based studies) [1], [2], [3], [4], [6]. Another group of studies utilize performance measures that are either based on quality, or quality-adjusted measures [7], [8], [9], [10], [12]. Only a few studies use several measures, based on different dimensions [4], [11].

Although trade-offs between dimensions (e.g. the quality and the quantity dimension) are not unlikely, there are often incentives in place to prevent workers from focusing their effort only at one of the dimensions. Both explicit incentives, for example by decreasing their output if workers are required to re-do a task in case of a defective product or service [1], and implicit incentives, for instance if supervisors can identify low-performing workers and are able to sanction them, help to mitigate problems related to multiple dimensions.

Data sources and aggregation

Most studies that use measures of worker productivity employ data from firms, e.g. from firms' internal databases, or public data, e.g. publication databases. These data allow analysis of workers' productivity on the individual level. For some jobs, though, workers' performance cannot be measured at the individual level, but only at a more aggregated level. Examples of this include firms that rely on team-based production, i.e. where output is jointly produced by a team of workers, or joint publications in academic research.

In some cases, the output of a whole group is of interest, rather than individual productivity. This is the case, for example, if one is interested in how incentives for team managers affect team performance. Alternatively, the interest may be in disentangling group-based measures into individual contributions to draw conclusions about individual behavior. However, it is difficult to reliably disentangle group performance into its individual contributions. If the group composition does not change, for example, it is empirically impossible to distinguish between the team members' individual contributions. If the group composition varies over

time though, e.g. when co-authorships change, one can estimate each individual's contribution to the team's overall output.

Time dimension

An advantage with direct, i.e. individual, measures of performance is that these are often available at a higher frequency compared to measures of aggregate productivity or wages. Figure 2 shows that some of the selected measures are available by the week, day, and in some cases, even by the second.

While this level of detail is not always necessary at the managerial level, it can inform about a range of research questions and thereby assist in the decision-making process. This allows for a greater ability to measure variation across time compared to studies that use longer data point frequencies. One example of variation across time comes from changes that are unrelated to the workers' effort, such as from changes in customer demand in service-sector jobs, or from weather conditions in agriculture. Depending on the setting, it might be important to take these effects into account, e.g. by controlling for the specific day of the week. The contrary can be the case if the production process is rather lumpy, i.e. if output occurs infrequently. In these cases, such as in the number of publications for scientists, it is important to set measurement periods that are sufficiently long.

How can measures of worker productivity be used to inform policy?

Studies using measures of worker productivity originated from performance analyses in private-sector firms, usually focusing on questions and decisions related to human resource (HR) management. These studies, which often make use of a single firm's personnel data, are referred to as "insider-econometrics studies." They exploit exogenous shocks to identify the causal effects of treatments on an outcome, using the performance measure as a metric [1], [4], [6]. Increased digitalization facilitates the gathering of performance data for both researchers and managers. In fact, firms collect performance measures not only to evaluate workers, but also to experiment and improve aspects related to HR management. A reliance on direct measures of productivity is by no means limited to the private sector. Measures of workers' productivity have been applied to a wide range of fields, such as research and education, health care, and politics (see Figure 2).

Using workers' productivity for setting incentives

A large number of studies use direct measures of workers' productivity in conjunction with either a change in the organization's incentive system or a randomized experiment to analyze how incentives affect individuals' behavior. The common finding is that monetary incentives affect workers' effort, resulting in higher performance. Besides the effect on workers' intrinsic motivation, incentives can also affect the sorting of workers, which can explain up to 50% of the gains in productivity [1].

These types of studies have also been applied to organizations whose primary goal is not that of profit maximization, such as schools, universities, and even politics. A number of studies, for example, use value-added measures for teachers to analyze their productivity. Measures of teacher value-added are based on an estimation of students' test scores, controlling for student and school characteristics, and capture the additional input a teacher has on the students'

grade outcomes [9], [10]. Studies have shown that this measure of teachers' productivity is indeed predictive of students' later outcomes, such as college attendance and salaries [9]. In US Charter Schools (schools that receive government funding but operate independently of the established public school system in which they are located), teacher value-added measures are increasingly used to evaluate and incentivize teachers.

More direct worker productivity measures are available in academic research as well as in firm based research and development. Scientists' productivity can be measured according to various metrics, such as patent statistics, (quality-adjusted) counts of publications, or citations [7], [8]. As is often the case in the private sector, the introduction of monetary incentives for publications is shown to increase research productivity.

Although less obvious, the same logic used to measure workers' productivity can be applied to almost any field that has some degree of quantifiable outcomes. One group of workers that is rarely related to productive behavior is politicians. But even for this group, it is possible to construct performance measures to demonstrate how, for example, remuneration of politicians can be optimally designed. One method to measure politicians' performance is the number and type of bills submitted [12].

An important element when defining incentive systems is the time frame over which performance is evaluated. Research has shown that if the date at which they are to be evaluated is close at hand, workers might increase their effort in order to achieve a bonus; or they might decrease effort if they are not able to achieve the bonus. Higher frequency performance measures, e.g. at the daily or weekly level, can help to understand these dynamic patterns in how individuals respond to incentives [13].

Formal and informal human capital acquisition in firms

Measures of worker productivity can also be used to estimate how workers accumulate human capital. Informal learning on-the-job as well as formal learning in training programs can increase workers' human capital, thus increasing individual productivity, according to performance measures.

This can be shown, for instance, by analyzing how job performance increases with job tenure. The common finding is that performance increases strongly in the early employment period, whereas the marginal return to tenure decreases over time [2]. Likewise, workers' productivity measures can be used to estimate the returns to training programs [4]. In both cases, direct measures of workers' productivity allow for an estimation of how learning affects an individual's job performance. In contrast, wages, which are often fixed in the short term, might not exhibit these effects. Further, wages would only capture returns to training if the benefits from increased productivity were shared between employer and employee.

Peer effects

Peer effects, i.e. changes in workers' productivity due to the presence of peers in the workplace are driven by two distinct mechanisms. They can either arise because of social pressure, or because of peer learning, i.e. spillover effects. Although human capital externalities have long been part of growth models, peer effects in the workplace have been analyzed only in recent years. This may be due, in part, to the availability of detailed information on workers' networks and detailed measures of their productivity. These studies often exploit settings in which workers are randomly exposed to high-productivity co-workers to estimate peer

effects in private-sector firms and schools [3], [4], [6], [10]. Peer effects can also be analyzed in knowledge-intensive environments, such as research. Very detailed measures on research productivity (e.g. publication output and citations), collected over a sufficiently long time, allow for the estimation of peer effects from high-productive researchers on their co-workers [8].

Working hours and performance

Another example of how direct measures of workers' productivity can inform policy involves working hours. The length of the standard workweek varies both between and within countries. Little is known about what an optimal number of weekly working hours would be from an efficiency perspective. Direct measures of workers' productivity allow one to estimate how the number of working hours or shifts in working hours affect performance; this might be crucial with respect to health and safety. Long working hours, for example, might result in increased fatigue, which is particularly dangerous in medical occupations [11]. This type of result can also provide direct advice about whether working hours should be expanded or reduced.

Gender pay gaps

One advantage with direct performance measures is that they are often objective and thus less likely to be affected by subjective evaluations. This objectivity can be especially useful when analyzing wage gaps between groups on the labor market, e.g. between men and women. Although gender pay gaps have decreased over time, women still earn considerably less than men. While there are many arguments for why gender pay gaps exist, it is important to understand to what degree the gender pay gap depends on underlying productivity differences. Direct measures of performance can help examine whether gender pay gaps are based on differences in productivity, how these gaps can be explained, and how possible gender differences in performance affect career choices. For lawyers, gender gaps in productivity are found to occur early in their careers, and have long-lasting effects on earnings and career advancement [5]. While objective measures of performance can help to identify gender gaps, it is more difficult to assess how accurately they reflect differential effort choices from men and women, e.g. if workers expect to be discriminated against.

LIMITATIONS AND GAPS

It is evident that there is no universal measure of workers' productivity. Rather, there are various measures that capture workers' performance in their specific settings. While this has the advantage of allowing highly detailed analysis of the determinants of workers' productivity, it comes at a cost.

First, even if performance can be measured for a wide range of occupations, the measures used are based on single workplaces, which are not representative of the whole sector or economy. While these data allow for comparisons between workers within the same tasks or occupation, it is rarely possible to make direct comparisons of workers' performance across different occupations. It is, however, possible to compare estimates based on several studies to draw more representative conclusions.

Second, the specificity of many performance measures makes comparisons over long time periods or across multiple countries difficult. While performance measures often allow for

the analysis of day-to-day or week-to-week variations of worker behavior, they are generally only available for rather short periods of time [4], [6]. Only a few studies exploit longer data sets covering several years, which allow for the analysis of trends [9]. For these reasons, direct measures of workers' productivity are not always appropriate. Instead, other measures, such as wages or firm value added may be more useful.

Lastly, although many occupations include some measures of performance, these often capture only one dimension. In some cases, e.g. when designing and evaluating incentive schemes, observing just one measure of performance might hide important aspects of workers' behavior, due to the multidimensional nature of productivity.

SUMMARY AND POLICY ADVICE

Direct measures of workers' productivity are not only used in personnel economics, but also in other fields, such as education and health, and are used in both the private and the public sectors. These measures are useful to inform policymakers about individuals' behavior at work, and to improve decision making, such as setting appropriate incentives.

Measures of workers' productivity represent a relatively new tool for economics research. Even if there is no universal definition of workers' productivity, studies using these measures have made important contributions to a wide range of fields. These contributions are not limited to low-skilled routine jobs where performance measurement might be easier. Rather, measures of workers' productivity are available for a wide range of jobs, including low- and high-skilled jobs in both the private and public sectors.

Although these studies are often based on single firms, which are not representative of a whole sector or economy, the detail at which one can observe workers' behavior permits the examination of questions that are difficult to address using survey or register data. Given the specificity of the settings, an appropriate way to achieve more "representative" results is to conduct studies with similar questions in different settings, e.g. in different firms. Meta studies can then be used to find a consensus, which could come close to the average effect for an economy.

Given the current state of research on using workers' productivity measures, policymakers and managers alike should select the right measures that will help them make informed decisions, for instance, when it comes to setting and designing incentives, or when regulating working hours.

Acknowledgments

The author thanks two anonymous referees, and the IZA World of Labor editors for many helpful suggestions on earlier drafts. The author also thanks Nora Döring.

Competing interests

The IZA World of Labor project is committed to the *IZA Guiding Principles of Research Integrity*. The author declares to have observed these principles.

© Jan Sauermann

REFERENCES

Further reading

Ichniowski, C., and K. L. Shaw. "Insider econometrics: Empirical studies of how management matters." In: Gibbons, R., and J. Roberts, *Handbook of Organizational Economics*. Princeton, NJ: Princeton University Press, 2013; pp. 263–311.

Shaw, K. L. "Insider econometrics: A roadmap with stops along the way." *Labour Economics* 16:6 (2009): 607–617.

Key references

- [1] Lazear, E. "Performance pay and productivity." *American Economic Review* 90:5 (2000): 1346–1361.
- [2] Shaw, K., and E. Lazear. "Tenure and output." *Labour Economics* 15:4 (2008): 704–723.
- [3] Mas, A., and E. Moretti. "Peers at work." *American Economic Review* 99:1 (2009): 112–145.
- [4] De Grip, A., and J. Sauermann. "The effects of training on own and co-worker productivity: Evidence from a field experiment." *The Economic Journal* 122:560 (2012): 376–399.
- [5] Azmat, G., and R. Ferrer. *Gender Gaps in Performance: Evidence from Young Lawyers*. IZA Discussion Paper No. 9417, 2015.
- [6] Bandiera, O., I. Barankay, and I. Rasul. "Social incentives in the workplace." *Review of Economic Studies* 77:2 (2010): 417–458.
- [7] Waldinger, F. "Quality matters: The expulsion of professors and the consequences for PhD student outcomes in Nazi Germany." *Journal of Political Economy* 118:4 (2010): 787–831.
- [8] Azoulay, P., J. Graff Zivin, and J. Wang. "Superstar extinction." *Quarterly Journal of Economics* 125:2 (2010): 549–589.
- [9] Chetty, R., J. N. Friedman, and J. E. Rockoff. "Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood." *The American Economic Review* 104:9 (2014): 2633–2679.
- [10] Jackson, C. K., and E. Bruegemann. "Teaching students and teaching each other: The importance of peer learning for teachers." *American Economic Journal: Applied Economics* 1:4 (2009): 85–108.
- [11] Brachet, T., G. David, and A. M. Drechsler. "The effect of shift structure on performance." *American Economic Journal: Applied Economics* 4:2 (2012): 219–246.
- [12] Ferraz, C., and F. Finan. *Motivating Politicians: The Impacts of Monetary Incentives on Quality and Performance*. NBER Working Papers No. 14906, 2009.
- [13] Asch, B. J. "Do incentives matter? The case of navy recruiters." *Industrial and Labor Relations Review* 43:3 (1990): 89S–106S.

Online extras

The **full reference list** for this article is available from:

<http://wol.iza.org/articles/performance-measures-and-worker-productivity>

View the **evidence map** for this article:

<http://wol.iza.org/articles/performance-measures-and-worker-productivity/map>