



4-25-2023

Measuring Performance: The Economics of Cleaning the Outside of the Cup

Joshua Hollinger

Follow this and additional works at: https://digitalcollections.dordt.edu/faculty_work



Part of the [Christianity Commons](#), and the [Economics Commons](#)

Measuring Performance: The Economics of Cleaning the Outside of the Cup

Abstract

"Sometimes focusing too much on what we can measure distorts our beliefs and actions as we turn away from immeasurable things that matter."

Posting about the economics of evaluating workers based on their performance from *In All Things* - an online journal for critical reflection on faith, culture, art, and every ordinary-yet-graced square inch of God's creation.

<https://inallthings.org/measuring-performance-the-economics-of-cleaning-the-outside-of-the-cup/>

Keywords

In All Things, performance, measurement, merit pay, education

Disciplines

Christianity | Economics

Comments

In All Things is a publication of the [Andreas Center for Reformed Scholarship and Service at Dordt University](#).

Measuring Performance: the Economics of Cleaning the Outside of the Cup

Joshua Hollinger

April 25, 2023

Why do we try to measure performance? As an economics professor, I'm expected to quantify each student's performance at the end of the semester. Corporations, sports teams, and universities increasingly rely on data to measure individual and group success. Educators are often evaluated based on the test scores of their students. Even in the church, spiritual leaders can be assessed by attendance numbers, book sales, or dollars raised. But when is trying to quantify our performance a good idea, and how should we use measurements wisely?

We measure performance to simplify the complex information resulting from our actions, empowering better feedback and decisions. These measurements are beneficial to the extent that they lead to actions that advance our true objectives. A measure must have sufficient *scope* and *precision* for its intended use. Or conversely, we can't use a measure in a way that surpasses its scope and precision. Sometimes focusing too much on what we can measure distorts our beliefs and actions as we turn away from immeasurable things that matter.

Wise use of performance measures requires an understanding of the workers being measured, the nature of the work, and the ultimate goals of the work. It also requires some economic thinking, as we consider the costs and benefits of using a performance measure in a particular setting and manner. In this article, I'll discuss the economics of evaluating workers based on their performance. Then I'll apply this to two areas of personal interest: measuring the effectiveness of teachers and measuring righteousness in the Christian life.

The Economics of Performance Pay

For most jobs, better performance comes with rewards: promotions, bonuses, praise, or status. To reward good performance it must be measured, either implicitly or explicitly. This is the obvious reason for measuring performance: to encourage better performance, either by screening for the best workers or incentivizing effort.¹ The best workers are more likely to choose an employer that rewards them more, and workers are more likely to do their best work when rewarded for it.

Economics is about allocating resources, including workers and their effort, so the economic question is how to set up a policy that does this optimally. One scheme employers can use is *performance pay*, where a worker's pay varies with a measure of their performance. In many jobs, the boss uses discretion to determine raises and promotions, implying an *implicit* measure of worker performance². Most jobs also come with an implicit punishment associated with low performance, namely the threat of getting fired. But with performance pay, the measure is *explicit*, as employers calculate pay as a function of worker output.

The economic theory of performance pay is simple in essence. The employer cares about the worker's output. The worker cares about their pay and dislikes having to put in extra effort. So paying the worker more if they have higher output can incentivize the worker to increase effort. Does this work in practice? Sometimes, yes. Ed Lazear wrote a seminal paper in 2000³ based on data for the Safelite Glass Corporation. Safelite installs glass windshields on cars. In 1994, they implemented performance pay, where workers were paid based on the number of units they installed rather than by the hour. Lazear found that the scheme drastically increased worker output and company profits. This also works well with fruit picking, as people pick more fruit when paid based on how much fruit they pick. For a similar reason, many salespeople are paid by commission: paying them based on how much they sell incentivizes more selling.

When It Works / When It Doesn't

Performance pay works well when the measure adequately captures the objective of the employer. Safelite wanted more windshields installed and that's what they got. But for many other workers, our performance measures are more limited in *scope* and *precision* relative to the ultimate objective of the work. *Scope* refers to how much of a worker's performance is measured. *Precision* refers to how accurately a measure captures the aspect(s) of performance it intends to capture. When a measure lacks sufficient scope, attaching an incentive to it might cause the worker to focus too much on measured performance at the expense of unmeasured performance.⁴ When a measure lacks precision, attaching an incentive to it leads to risk for the worker, unclear feedback, and unfair rewards.

Performance Pay in Education

In my own research, I've been studying incentives in K-12 education. Education plays an important role in providing skills to future workers and a chance at upward mobility for kids coming from poorer backgrounds. Market forces usually encourage companies to pursue profits, which gives employers a reason to incentivize high performance from their workers. But these forces mostly don't apply in public education, where there's often little competition between schools. A compressed pay schedule, where teachers are typically all paid the same except for increases based on experience or an advanced degree, may leave teachers with little incentive to perform well. Around 90% of students attend public schools. Since the government

is responsible for running these schools well, a long-standing public policy question has been how to use incentives in education.

An incentive scheme can use “the carrot” or “the stick”. By the turn of the 21st century, a bi-partisan consensus emerged on “the stick” approach, often called “school accountability”. This approach went nationwide with the passing of the No Child Left Behind Act (NCLB) in 2001, and it focused heavily on student test scores as a measure of school performance. Schools continuing to have too many students testing below the proficiency level faced escalating sanctions. Several studies have concluded that the policy was at least somewhat successful in raising test scores.⁵ But test scores might not capture the full range of skills and knowledge students acquire through education, leading to the concern of “teaching to the test”. What if, in response to strong incentives to increase test scores, schools focus too narrowly on test preparation at the expense of other important aspects of learning?

Several studies have documented unintended responses to NCLB, such as schools focusing more on tested subjects rather than untested subjects.⁶ or increasing the calorie content of school lunches on test days.⁷ But I wanted to look at the long-run effects of the policy to see how big of a problem “teaching to the test” might be.⁸ I found that NCLB increased elementary students’ test scores, and years later, these students performed better on tests like the SAT and ACT. But there was no effect on high school graduation or college-going. An increase in math and reading skills may have been valuable, but I show evidence that schools shifted their efforts toward better performance on tests. Thankfully I didn’t find any negative effects on other long-run outcomes, but the policy seems to have been limited by the scope of how it measured school performance.

Precision

There were plenty of problems with the design of NCLB. One is that it largely punished schools for having low-performing students, not for *causing* low performance on tests. As NCLB came to an end, the emphasis on test scores decreased, and states were given much more discretion in designing accountability policies. Several states and districts started using “the carrot” approach, paying teachers bonuses if their students exhibited high test score gains. This comes closer to capturing the causal effect of teachers, because it controls for their students’ test scores in the prior year. But we run into another problem: elementary teachers often only have around 25 students per year. Trying to measure a teacher’s performance in a given year is difficult, because the sample size is too small. The test score gains of 25 students will naturally fluctuate due to a lot of factors that have nothing to do with the teacher. In another paper, I studied one of these performance pay programs and found that the imprecision of the single-year performance could explain why teachers didn’t respond much to the program.⁹ Anecdotally, many teachers felt the bonuses paid were quite random, having little to do with teacher performance. Even if all we cared about was test scores, the measure wasn’t precise enough to be used in that way. A longer-term approach that considers multiple years of data and other factors, such as classroom observations and peer evaluations, may be more

precise and effective in evaluating teacher effectiveness. Recent evidence suggests performance pay for teachers can work if done well.¹⁰

Christian Living

In the Christian life, the temptation to focus on external measures of faithfulness can lead to a neglect of the internal attitudes and motivations that truly reflect a transformed heart. In Matthew 23, Jesus rebukes the religious leaders of his day for their emphasis on external measures of righteousness, such as their outward displays of piety or strict adherence to man-made rules, while neglecting the internal matters of the heart:

“Woe to you, teachers of the law and Pharisees, you hypocrites! You clean the outside of the cup and dish, but inside they are full of greed and self-indulgence. Blind Pharisee! First clean the inside of the cup and dish, and then the outside also will be clean.” (Matthew 23:25-26)

We can see the “outside of the cup”. But our measurement of how other Christians are performing lacks scope and precision. Legalism and related issues arise if we over-emphasize the things we can measure, like not mowing lawns on Sundays, expressive public prayers, participation in Christian groups, or posting our devotions on social media. None of these things are necessarily wrong, and we need some practical guidelines for walking in righteousness individually and communally. We are called to hold each other accountable. But we must be careful not to judge others in a way that oversteps our very limited knowledge of their hearts. If we focus too much on what we can measure, we encourage hypocrisy, as people clean the outside of the cup at the expense of transforming the inside of the cup. Ultimately, we’re all accountable to God, who sees everything perfectly.

Using Performance Measures Well

Measuring and rewarding performance can encourage us to improve. But we need to be aware of the limitations of what we can measure and guard against relying too heavily on insufficient measures. By doing so, we can work together now amidst our ignorance, weaknesses, and limitations, while looking forward to the day when God will perfectly reward us for the work He enables us to do.

-
1. If workers are good at different tasks, assessing their performance can also allow an employer to match workers to positions more efficiently.
 2. The boss may or may not use data in this decision, but choosing one worker over another for a promotion is logically equivalent to rating workers on a scale of 1 to

10 and choosing the worker with the higher rating. Thus, there's an implicit measure of performance.

3. Lazear, Edward, P. 2000. "Performance Pay and Productivity." *American Economic Review*, 90 (5): 1346-1361. DOI: 10.1257/aer.90.5.1346.
4. One useful model for understanding this issue is the multitasking model proposed by Holmstrom and Milgrom (1991) Holmstrom, Bengt, and Paul Milgrom (1991). "Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design." *Journal of Law, Economics, and Organization* 7: 24-52.
5. For example, see Chakrabarti, Rajashri (2014). "Incentives and responses under No Child Left Behind: Credible threats and the role of competition." *Journal of Public Economics*, 110, 124-146. Or Dee, Thomas S. and Brian Jacob (2011). "The impact of No Child Left Behind on student achievement." *Journal of Policy Analysis and Management*, 30, 418-446.
6. Jacob (2005) Jacob, Brian A. (2005). "Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago Public Schools." *Journal of Public Economics*, 89, 761-796.
7. Figlio and Winicki (2005) Figlio, David N. & Winicki, Joshua, 2005. "Food for thought: the effects of school accountability plans on school nutrition," *Journal of Public Economics*, Elsevier, vol. 89(2-3), pages 381-394, February.
8. Hollinger, Joshua, "School Accountability, Test Scores, and Long-Run Outcomes," 2021.
9. Hollinger, Joshua, "Performance Pay and Incentive Strength for Better and Worse Teachers," 2022.
10. See Cohodes, S., Eren, O., & Ozturk, O. (2023). *Teacher Performance Pay, Coaching, and Long-Run Student Outcomes* (No. w31056). National Bureau of Economic Research. And Morgan, A. J., Nguyen, M., Hanushek, E. A., Ost, B., & Rivkin, S. G.

(2023). Attracting and Retaining Highly Effective Educators in Hard-to-Staff Schools

(No. w31051). National Bureau of Economic Research.