Media and motivation: the effect of performance pay on writers and content

Ivan Balbuzanov†  Jared Gars‡  Emilia Tjernström‡

March 2, 2018

JOB MARKET PAPER
Preliminary and incomplete. Please do not cite.

Abstract

We study the causal effect of journalist incentives on the quantity, quality, and composition of media content. Using a field experiment that randomly allocated writers either to a contract paying a flat rate per article or to a contract that rewards writers based on pageviews, we demonstrate that output-based performance pay reduces the number of submitted articles, modestly increases per-article effort, and sharply increases pageviews – findings that are in line with our theoretical predictions. Specifically, the output-based contract results in a doubling of total pageviews, a 150% increase in pageviews per article, and a sharp reduction in the firm’s cost per “click.” Preliminary results suggest that the treatment has limited effects on quality or the prevalence of clickbait, but ongoing work employs supervised machine learning to better understand the impacts on content and quality. The experiment takes place within an online news firm in Kenya, where writers operate in a complex environment, choosing how many tasks to undertake and how to allocate effort across them. Our study suggests that output-based incentive contracts have substantial implications for journalists’ effort and content choices, with short- and long-term implications for firm profitability and for the quality of news.

*We thank Pascaline Dupas, Leah Lakdawala, Huailu Li, Teresa Molina, Laura Schechter, Alessandro Tarozzi, Maisy Wong and participants at the WGAPE 2017 annual meeting, Advances with Field Experiments Conference, Universitat Pompeu Fabra, and UW-Madison for helpful comments and discussions. This study is registered in the AEA RCT Registry and the unique identifying number is: AEARCTR-0002396. This work is supported by funding from the IFREE small grants program. The opinions expressed herein do not reflect the opinions of the sponsoring agencies. All remaining errors are our own.
†University of Melbourne, AUS; ivan.balbuzanov@unimelb.edu.au
‡University of Wisconsin-Madison, USA; gars@wisc.edu, tjernstroem@wisc.edu
1 Introduction

The news media plays an important role in the functioning of democratic economies, both by providing consumers with information about policies that affect individuals and firms, and by establishing citizen-government accountability relationships. These crucial roles of the media are especially important in developing countries, where weak institutions often leave large gaps in political accountability. While a growing literature investigates the effect of political and economic forces on media content at the newspaper level (see for example Beattie et al. 2017 and Di Tella and Franceschelli 2011), little empirical evidence exists on the relationship between individual journalists’ incentives and the quality of the news articles that they produce.

As performance contracts (also known as “pay-per-click” contracts) for online journalists have become more prevalent, the topic of journalists’ incentives has also received more attention (Murtha, 2015).1 Supporters of the model argue that readers value investigative journalism, and that giving writers greater discretion over how to connect with their readers should reward in-depth reporting. For example, Vogt et al. (2016) document that crowd-funded journalism projects have been increasing rapidly in recent years. In 2015, over 25,000 individuals contributed to such campaigns, and longer-form journalism stood out as the most commonly funded category, accounting for around 43% of crowd-funded projects. Detractors counter by noting that journalists might instead choose to “dumb down” stories, increasing their use of easy topics and “click-bait” headlines. Despite growing concerns about the impacts of changing business models on the quality of journalism and on the polarization of political news, “pay-per-click” contracts have received limited attention from economists. To our knowledge, ours is the first study to rigorously explore how output-based performance contracts affect journalistic quality and firm profits in the market for news.

Specifically, we employ a within-firm experimental design to study whether pay-per-click contracts affect the quantity and quality of news content. The experiment took place within an online news firm that sources all its news stories from local, independent reporters. Writers therefore operate as independent “gig-workers” who can choose how much to work, what to write about, and how much effort to allocate to each story. In an effort to reward higher-impact journalism

---

1For example, media firms such as Slant and Politico have implemented bonuses for every 500 clicks on top of a fixed wage; the online publishing platform Medium now rewards writers based on the “engagement” that their articles receive.
and to decrease the number of articles that receive unprofitably few views, writers were randomly allocated to contract groups: the first group (Control) continued receiving the status quo contract, consisting of a piece-rate per published article; the second group were transferred to a pay-per-view (PPV) contract with a piecewise linear pay structure in the number of unique views that the article received; the third group were allowed to choose between the two contracts (Opt-in). We find that the new contracts reduce the number of article submissions, modestly increase per-article effort, and sharply increase pageviews.

Estimating the intention-to-treat effect, we show that the (PPV) contract decreased treatment writers’ article submissions by 26 articles per month relative to the control group (mean = 49). Alongside the decrease in submissions, the output-based performance contract increased average views per article by 155 percent and doubled treatment writers’ total views, relative to the flat rate. These treatment effects were immediate and quite persistent. Taken together, these results suggest that writers had knowledge about some aspects of the production function but were not sufficiently incentivized to allocate effort or to use the contextual information under the piece rate contract.

The above results focus on the treatment effect of the new contracts on article popularity and writer effort, but we also care about the quality of the articles that writers produce. Preliminary results suggest that the treatment has limited effects on article quality (as rated by the firm’s editors) or on the prevalence of clickbait titles. We are in the process of generating data to gain a deeper understanding of the impacts on content and quality: we are currently recruiting workers from an online labor market to rate the articles on quality and on a two-dimensional political bias spectrum. We will then train a machine learning algorithm to predict quality and bias for the remainder of sample articles and estimate the effects of incentives on quality and bias.

Within the context of the broader performance pay literature, digital news constitutes an ideal context for studying output-based performance contracts: productivity (i.e., views) is relatively easy to measure, inputs are not easily defined or observed, and writers may have incomplete knowledge of the marginal returns to different actions. Our first contribution to the literature on incentive schemes in firms is to provide the first experimental evaluation of a switch from a flat piece rate to output-based incentive contracts for a multi-dimensional task with stochastic output. Previous empirical studies of performance contracts within firms focus primarily on a switch from hourly
wages to piece rates. By their nature, piece-rate contracts are common when tasks are simple or repetitive, inputs can be monitored (imperfectly), there is little scope for innovation or discretion, and output is not stochastic.\(^2\) Productivity impacts from a switch to piece rates range between 20 and 50 percent, and the impacts on profits tend to be dampened by increased managerial costs. Given that we find substantially higher productivity increases moving from a piece rate to an output-based contract, we interpret the difference as a result of journalists having discretion over their actions (delegation) and private information about the returns to actions that are insufficiently incentivized under the piece-rate contract (Prendergast, 2002; Raith, 2008).

The evidence on output-based pay complements existing experimental research on how to best incentivize public service delivery in developing countries. Mohanan et al. (2017) compare input- and output-incentive contracts for healthcare providers’ in Karnataka, India. They find that only healthcare providers with advanced qualifications reduced post-partum hemorrhage under the output-based contract, presumably because they could apply their knowledge and implement context-specific strategies to prevent bleeding. Muralidharan and Sundararaman (2011) focus on performance bonuses for teachers based on student test scores in Andhra Pradesh, India and find that bonuses lead to significant increases in student performance (0.27 SD increase in math test scores). Using teacher surveys, they find that teachers innovated after the introduction of the performance bonuses by assigning more homework and giving practice tests, and allocating effort towards weaker students. In both cases, the benefits of output-based pay are higher for multi-dimensional tasks when the production function is partially known by the principal, as is the case in the production of news.

We are also better positioned to examine the effects of performance pay on firm profits than most previous studies. While output-based contracts have been implemented in a variety of settings (Lazear, 2000; Finan et al., 2015) to overcome principal-agent problems, the empirical evidence on whether it actually matters for overall firm profits is inconclusive (Prendergast, 2015).\(^3\) While agents face more risk in uncertain environments, output-based contracts may be more efficient when agents have specific knowledge about the production function (Raith, 2008), or can learn the returns

---

\(^2\) For example, Lazear (2000) studies workers in a firm that install windshields, Shearer (2004) studies tree planters in Canada, and Bandiera et al. (2005) study fruit-pickers in the UK.

\(^3\) See Bandiera et al. (2011) or Miller and Babiarz (2013) for recent overviews of empirical evidence on pay-for-performance contracting in developing countries.
to specific actions. Further, it has been recognized that the marginal returns to delegation are higher when the principal cannot observe or define effort (Prendergast, 2002). However, contracting on outcomes may crowd out intrinsic motivation (Bénabou and Tirole, 2006), incentivize short-termist behavior that rewards only current output but may erode long term productivity (Bolton et al., 2006), and lead to a reduction in effort allocated to non-contracted outcomes (Holmstrom and Milgrom, 1991). We take advantage of the fact that firm profits in our setting are roughly proportional to pageviews, and estimate the effects of contract structure on firm profits.

Another key question in the literature has been whether performance contracts primarily impact workers through behavioral change or through selection effects (Bandiera et al., 2011). However, empirical evidence separating these effects in personnel settings is scarce. Lazear (2000) observes an individual firm with a staggered rollout of a piece-rate contract and finds that half of the 44 percent increase in productivity was due to more productive workers selecting in, i.e., joining the firm. However, as Dohmen and Falk (2011) point out, sorting may also take place along other dimensions than quality, potentially affecting the workforce composition in unanticipated ways. We therefore also examine the relative importance of incentives versus selection effects in driving the treatment effects that we observe. We are able to study selection in two ways: first, we compare the characteristics of writers who remain active after the contractual change across control and PPV. Secondly, our third treatment group (Opt-in) allows us a unique window into selection, which we exploit by comparing the characteristics of writers who choose pay-per-view contracts to those who select to remain in control.

Finally, we can test whether performance contracts increase experimentation, i.e. the acquisition and productivity-increasing use of information about the marginal returns to writer effort. While some theoretical work focuses on contracting when the marginal productivity of effort is stochastic (Zabojnik, 1996; Rantakari and others, 2008), empirical tests separating the information effect from the incentive effect of performance contracts remain limited. We explore a number of mechanisms to determine the margins over which writers change their behavior in response to the PPV contract. Writers can choose a number of actions for each article when deciding what to write, including the topic, the target audience, the amount of time spent on each article, and the characteristics of the

---

4 Relatedly, Bandiera et al. (2007) find that productivity increases when managers’ pay becomes more closely tied to output, and attribute roughly half of these worker-level productivity increases to manager discrimination in hiring.
written language. We show that active writers respond to the PPV treatment by increasing the share of political articles by 26 percent (from a baseline of 33.4 percent), and that they contribute \textbf{X percent} fewer articles on local (county-level) topics as a share of submissions.\(^5\) This increase in per-article impact is present regardless of topic or location choice. Finally, self-reported effort per article increased by a modest 4 minutes per article from a baseline of roughly 32 (not statistically distinguishable from zero), and drop by almost 460 minutes/month (mean = 1,351).

The paper proceeds as follows. Section 2 provides information on the online news firm and the market structure in which it operates. Section 3 describes the experimental design, the randomization, and how the firm calibrated the new contract. Section 4 develops a theoretical model that explores how the new output-based performance contracts affect writer behavior (per-article effort and submission frequency) and welfare. Finally, Section 5 describes the various data sources that we draw upon, while Section 6 provides both graphical and econometric evidence for our main research questions, and additionally explores some of the mechanisms behind these impacts. Section 7 explores the robustness of the results, and Section 8 concludes with a discussion of the broader implications and future work.

2 Background and Setting

2.1 Media market

Kenya’s post-independence media landscape enjoys a fairly broad range of media outlets, with a mix of state-owned and private media outlets, most of which publish in English (Lohner et al., 2016; Freedom House, 2017). The news media consist of at least four daily newspapers, one business daily, and a handful of regional weekly newspapers (Freedom House, 2017). Although media coverage has traditionally been relatively rigorous, especially in the print sector, editorial pressure and the political preferences of advertising sources shape coverage at many outlets (Freedom House, 2017). Furthermore, reporters can face repercussions for critical coverage, including dismissal.

Additionally, human rights groups have expressed concerns in recent years that press freedom is in sharp decline (Namwaya, 2018; Freedom House, 2017).\(^6\) Some sources associate the reduction

\(^5\)County-level articles naturally have a smaller potential audience than national-level topics.

\(^6\)The most recent example occurred in early 2018, when the Kenya Communications Authority switched off a number of television stations for broadcasting live from the site of a political action by the opposition leader. The
in press freedom directly with ownership concentration and media capture. Freedom House (2017) notes that most media outlets are owned by politically connected people or politicians, and that five media companies capture over 70 percent of all media consumers. The public’s trust in the media sector is, perhaps unsurprisingly, low. Simiyu (2014) surveys two small but broadly representative samples of citizens and reporters after the 2013 presidential elections, and finds that a stark majority of the Kenyan electorate believe the media was biased and partisan in their reporting of the election. The journalists sampled The journalists corroborated these perceptions, stating that their journalistic freedom was tightly limited during the coverage of the election.

While the US media is a great deal freer than its Kenyan counterpart, there are some interesting parallels. In particular, both media markets are characterized by high ownership concentration and low citizen confidence in mass media. Ladd (2011) reports that the portion of Americans expressing “hardly any” confidence in the press had risen to 45% by 2008, and at that time a record high of 63 percent agreed with the view that reporters were politically biased in their reporting. More recent Gallup polls ("Gallup Poll", 2016) show that Americans’ trust and confidence in the media has continued to set records: ”to report the news fully, accurately and fairly” has dropped to its lowest level in Gallup polling history, with only 32% saying they have a great deal or fair amount of trust in the media. US media industry concentration rose over the period 2004-2013, across various indices. Noam (2016) notes that cross-ownership tends to be high both at the lower and higher ends of the economic development spectrum. The main difference is that government ownership tends to play a larger role in less-developed countries than in highly developed media markets. The US appears in the top-five list of highest cross-ownership percentages (40.4%), together with China (73.9 %), Sweden (44.4 %), Brazil (40.9 %) and Canada (35.9 %). As global news consumption moves increasingly online (including social media), understanding the response of “gig-worker” journalists to changing pay structures seems of high relevance to current policy debates.

2.2 The news firm

We study an online newspaper in Kenya that sources its stories from local reporters, who have been paid a piece-rate for each article they publish since the website’s launch in 2014. Writers need...
no previous qualifications to submit content, only a device and basic writing skills. The website averaged over 1.5 million active readers between May 1st and September 1st of 2017 (the four months leading up to the experiment) and has experienced dramatic growth: average monthly readers grew from 68,000 in January 2016 to 887,000 in December that year. As can be seen in Figure 1, which depicts daily pageviews since 2014, views have declined since the end of the election period, but the overall trend since 2016 is still upward.

After signing up, writers can submit articles for publication subject to a 500-word limit. Each writer has a “workbench” page where they can edit their profile, review and submit articles, and see their payment history. In addition to their workbench, each writer has a public profile that links to their published articles, listing the date of publication and the number of pageviews that the article has received. It is entirely up to writers what stories they want to cover, and how frequently they want to submit. Professional editors curate all incoming articles and choose which ones to publish, monitoring for defamation, plagiarism, hate speech, and a few more issues that guarantee rejection.

Articles are reviewed by editors in the order in which they are submitted. While the firm encourages timely submission of articles on breaking news events, articles must meet a minimum

---

8The sharpest increase in pageviews coincides with the run-up to the Kenyan presidential election on August 8, 2017 (Election #1), where the incumbent President Uhuru Kenyatta won reelection with 54% of the vote. The election results were challenged in the Supreme Court by the opponent, Raila Odinga, and the result was annulled under the condition that fresh elections would be held on October 17, 2017 (Election #2).
quality standard to be published. The primary reasons for rejection, in the order of frequency, are for being too similar to published articles, not engaging enough, plagiarized, and too informal. There is no editorial discretion in the topics of articles that get published and neither the editors nor the firm suggest topics or content. Prior to the contract change, writers were paid 100 KES (about 1 USD) via the mobile money system M-Pesa if an article was published.9

In the two months leading up to the contract change, 122 writers submitted articles and 107 successfully published at least one article. This resulted in an average of 830 published articles per week. Of the two months, these roughly hundred writers were active for 4.3 weeks on average, where we define active as submitting at least one article in a given week. Conditional on being active, the average weekly number of submissions and publications were 7.67 and 6.64 articles, respectively. However, there is substantial heterogeneity in the activity across writers in the numbers and rates of submission and publication. The median writer published only 2.35 articles per week while the top quartile published almost 13 articles. Further, 40 percent of writers submitted in 3 or fewer weeks out of the two months, and only 30 percent submitted in all 8 weeks. Half of all articles published over the period were submitted by just 13 writers and one third of writers contributed five articles or less. This suggests overall that the “gig workers” that we study range from depending quite heavily on writing as a source of income, to very occasional writers. As described in more detail in Section 3, we stratify the randomization on various measures of participation to ensure balance across these different types of writers.

3 Experimental design

To examine the effects of an output based incentive contract for writers on the quality and quantity of their output, we randomly varied writers’ contractual terms for the treatment groups while maintaining the flat rate (status-quo) contract for the control group. Section 3.1 describes these treatments in more detail. Section 3.2 then describes the randomization, while Section 3.3 discusses how the performance pay contract was calibrated.

Payment for published articles was disbursed weekly until June 6, 2017. Since then, writers have been paid every three days.

---

9Payment for published articles was disbursed weekly until June 6, 2017. Since then, writers have been paid every three days.
3.1 Experimental treatments

All writers who had published an article in the last 12 months since September 5, 2017 were randomly allocated into one of the following three treatment groups:

- **Control group/ Piece-rate:** The control group contracts maintain the status-quo contract of 100 KES (1 USD) per published article. This contract has been in place since the inception of the website in May of 2014. To receive payment in the control group, articles must pass the basic editorial bar (no repeat articles, no plagiarism, no hate speech, no defamation) and writers are paid within three days of publication.

- **Pay-per-view (PPV):** The pay-per-view treatment group contract pays out based on a kinked linear fee structure that is a function of the number of unique views that the article generates.\(^\text{10}\) The kinked fee structure is defined as follows: (i) there is a lower threshold of 400 views, below which writers receive no pay at all; (ii) between 400 and 800 views, writers receive a rate of 125 KES per 1000 views; and (iii) all pageviews beyond 1000, writers receive a lower rate of 12 KES per 1000 views, with no upper limit. Section ?? provides more details on how the contract was calibrated. Payments are issued within three days of publication based on the number of users that have accrued at the time that payroll is processed. Writers can continue to be paid for an individual article in every following pay period if the article continues to attract users.\(^\text{11}\) This fee structure applies to all users after publication of the article and does not restart in each pay period.

- **Contract choice (Choice):** The third group were allowed to choose between the status quo piece-rate and PPV. They were required to select their contract when they logged in to submit an articles as soon as the intervention had launched. Writers in the choice group were informed that their selection would be permanent. The terms and payment timing of the two options are identical to those described above.

The firm informed writers of the new contracts via SMS on the day of the intervention launch. All writers received a message from the Head Editor on their workbench that updated them on

\(^\text{10}\)The exact calculation is based on the “users” metric in Google Analytics, which measures the number of *unique* views by IP address. This metric prevents writers from easily manipulating their payments by, for example, refreshing the page in their browser. We will use “users” and “views” interchangeably throughout the paper.

\(^\text{11}\)Ninety-seven percent of pageviews are received within the first two days of publication on average.
the success and growth of the firm and thanked them for their participation in that success. This message was partly to ensure that any potential prompting effects of a message from the editor would be held equal across treatment and control, as well as to update all writers’ current knowledge about firm growth. Writers in the treatment groups were told, in addition to the basic message, that they were to be paid according to a new contract, and that the change would take place immediately. The language to PPV writers included a message explaining that the firm was trying to find a better way to reward successful writers for their work. The writers were not shown the complete pay schedule, but rather a table noting the key kinks in the payout schedule (as seen in Figure 2).

![Figure 2: Payment schedule](image)

PPV writers also received some information about the data that their payments would be based on (values from Google Analytics), about the very high correlation between “users” and the pageviews that they could already see on their profile pages, and about where they could locate their payment information for each article. Writers in the choice group received the same message, but were told that they could choose their contract by logging into their writer page and selecting their preference.

The compensation and responsibilities of the firm’s editors remained unchanged during the duration of the experiment and they were blind to the treatment status of writers, to preclude any confounding the effects of treatment with editor behavior. Editors are paid a fixed hourly wage for their editorial work but are also permitted to submit articles for publication. Editors that continued to submit articles were paid under the status-quo contract.
3.2 Randomization

All writers that had published at least one article in the year prior to the treatment (i.e., since September 5, 2016) were included in the randomization. In order to ensure balance on key baseline characteristics writers were stratified into 8 groups based on whether they had published at least one article in the month prior to treatment (0/1), if they submitted more than one article in the weeks when they were active (0/1), and if they were above or below median views per article for the period that they were active (0/1).\textsuperscript{12} Section Table 1 shows balance tests. Writers assigned to Control, PPV and Choice treatments are well balanced along observables at baseline, including how long since they first signed up (in weeks), how many of those weeks they were active, the number of submitted and published articles, and their average pageviews (both per-week and per-article). Of the 480 writers included in the sample, twenty-five percent (172) published at least one article within two months prior to the intervention and forty percent (235) published within four months prior.

Any new writers who signed up to write for the firm following the introduction of the treatment were also randomly assigned to a treatment group after registering as a writer. Advertisements on the website do not include information about the treatment or the pay structure and potential writers would only be aware of the available contracts if they knew one of the current writers. We attempted to prevent writers from registering again to get a more preferred contract by requiring that no email address can be registered twice and that no phone number can receive payment for more than one email address, but this enforcement was not perfect. Section 7 discusses this in more detail.

3.3 Contract parameters and design

Figure 3 shows the actual payoff schedule that writers faced. The firm calibrated these contract parameters based on the previous months’ user data, with the dual aims of incentivizing high-impact articles, while ensuring that the \textit{ex ante} expected payout would remain the same.\textsuperscript{13}

\textsuperscript{12}The views per article were time-demeaned to account for the growth in views over time; we worried that a non-demeaned views/article measure would introduce a correlation between views and the amount of time that had passed since the writer was last active.

\textsuperscript{13}Given the stochastic nature of output-based pay in the market for news, the literature suggests that risk averse writers may need to be compensated for the added risk. However, empirical evidence on the incentives-risk tradeoff remains scarce, and the relation between risk and performance contracting may be “too subtle” to consider in the
Table 1: Balance at baseline across treatments

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5) vs. (2)</th>
<th>(1) vs. (3)</th>
<th>(2) vs. (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tenure</td>
<td>34.867</td>
<td>35.484</td>
<td>35.509</td>
<td>35.290</td>
<td>0.727</td>
<td>0.722</td>
<td>0.989</td>
</tr>
<tr>
<td></td>
<td>(1.304)</td>
<td>(1.198)</td>
<td>(1.247)</td>
<td>(0.720)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weeks active</td>
<td>11.943</td>
<td>11.621</td>
<td>13.379</td>
<td>12.317</td>
<td>0.841</td>
<td>0.380</td>
<td>0.288</td>
</tr>
<tr>
<td></td>
<td>(1.123)</td>
<td>(1.149)</td>
<td>(1.185)</td>
<td>(0.665)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recent=1</td>
<td>0.266</td>
<td>0.261</td>
<td>0.267</td>
<td>0.265</td>
<td>0.920</td>
<td>0.980</td>
<td>0.900</td>
</tr>
<tr>
<td></td>
<td>(0.035)</td>
<td>(0.035)</td>
<td>(0.035)</td>
<td>(0.020)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Submitted</td>
<td>1.627</td>
<td>2.085</td>
<td>2.270</td>
<td>1.996</td>
<td>0.500</td>
<td>0.341</td>
<td>0.805</td>
</tr>
<tr>
<td></td>
<td>(0.421)</td>
<td>(0.532)</td>
<td>(0.525)</td>
<td>(0.286)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Published</td>
<td>1.410</td>
<td>1.848</td>
<td>1.943</td>
<td>1.735</td>
<td>0.482</td>
<td>0.392</td>
<td>0.891</td>
</tr>
<tr>
<td></td>
<td>(0.383)</td>
<td>(0.488)</td>
<td>(0.488)</td>
<td>(0.263)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Users (week)</td>
<td>2871.862</td>
<td>4789.852</td>
<td>5989.536</td>
<td>4560.908</td>
<td>0.368</td>
<td>0.218</td>
<td>0.691</td>
</tr>
<tr>
<td></td>
<td>(915.862)</td>
<td>(1908.697)</td>
<td>(2333.400)</td>
<td>(1054.572)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Users (article)</td>
<td>1851.533</td>
<td>1924.495</td>
<td>1839.807</td>
<td>1872.126</td>
<td>0.878</td>
<td>0.979</td>
<td>0.847</td>
</tr>
<tr>
<td></td>
<td>(341.789)</td>
<td>(329.345)</td>
<td>(285.980)</td>
<td>(182.875)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Share politics</td>
<td>0.316</td>
<td>0.302</td>
<td>0.355</td>
<td>0.325</td>
<td>0.692</td>
<td>0.306</td>
<td>0.154</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.026)</td>
<td>(0.027)</td>
<td>(0.015)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Share national</td>
<td>0.717</td>
<td>0.739</td>
<td>0.741</td>
<td>0.732</td>
<td>0.599</td>
<td>0.556</td>
<td>0.942</td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td>(0.027)</td>
<td>(0.028)</td>
<td>(0.016)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time (week)</td>
<td>449.127</td>
<td>237.540</td>
<td>329.441</td>
<td>338.012</td>
<td>0.347</td>
<td>0.603</td>
<td>0.383</td>
</tr>
<tr>
<td></td>
<td>(216.792)</td>
<td>(65.700)</td>
<td>(82.138)</td>
<td>(79.535)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time (article)</td>
<td>43.851</td>
<td>37.388</td>
<td>42.526</td>
<td>41.381</td>
<td>0.475</td>
<td>0.881</td>
<td>0.556</td>
</tr>
<tr>
<td></td>
<td>(6.473)</td>
<td>(6.234)</td>
<td>(5.935)</td>
<td>(3.564)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

N: 158 161 161 480

Standard errors in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01

The firm was also concerned with limiting the proportion of “losers” among current writers, which was the reason to introduce a steeper slope between 400 and 800 views. Of the articles used in the calibration, the firm chose values were such that 44 percent of articles in the PPV contract would earn less than the 100 KES they would have earned in the status-quo contract (the firm’s definition of “losers”) and 27 percent of articles would have received no payment at all. Meanwhile, nearly 11 percent of articles would have received double the realized status quo earnings of 100 KES. Thus, writers in the PPV treatment face substantial downside risk conditional on an equal amount of effort, but the returns to increased effort have the potential to outweigh these losses. It is also clear from Figure 3 that the non-linearity introduced by the kinks suggest that risk neutral writers, assuming that they have close control over the number of pageviews that their articles design of actual compensation plans (Raith, 2008).
recieve, should target producing roughly 1,000-view articles, since a single article needs to reach 9,500 articles to receive the same level of payout. As we will see, writers seem to have some control over their pageviews but this control is quite imprecise and they have imperfect knowledge about exactly how many pageviews their articles will receive.

![Figure 3: Payment schedule](image)

4 Theory

In this section, we present a one-period model to show how performance pay affects how writers choose the number of articles that they submit and how much effort to allocate across articles in a setting where output is stochastic. Assume that the writer chooses how many articles to write and submit \((n \in N)\) and how much effort to expend on each article on average \((e \geq 0)\). The writer’s utility equals the expected monetary payment minus the cost of effort. The cost-of-effort function \(c(e, n)\) is strictly convex in both variables and \(\frac{\partial^2 c(e, n)}{\partial n \partial e} > 0\) for all \(e, n\).

Conditional on the chosen effort level \(e\), the effective quality of each article is random and equals \(e + \epsilon\), where \(\epsilon\) is a continuous random variable with mean 0. If \(e + \epsilon < 0\), the article does not meet editorial standards, effort is wasted, and the writer receives no payment regardless of the chosen contract. If \(e + \epsilon > 0\), the article is approved and we assume that its quality, \(e + \epsilon\), is proportional...
to the number of pageviews it will attract.\(^{14}\)

Every payment contract we consider can be described by a pair of parameters \((\alpha, \beta)\), where \(\alpha\) is the fee paid per approved article and \(\beta\) is the scaling parameter that determines the quality-contingent payment. Thus, if \(e + \varepsilon > 0\), the writer’s payment is \(\alpha + \beta(e + \varepsilon)\).

We now separately consider the writer’s optimization problem when faced with a flat contract (FC), where \(\alpha_{FC} > 0\) but \(\beta_{FC} = 0\), and a linear (fee-per-user) contract (LC), where \(\alpha_{LC} = 0\) but \(\beta_{LC} > 0\).\(^{15}\) Under the flat contract, the writer chooses \(e\) and \(n\) to maximize her expected utility \(\mathbb{E}U_{FC}(e, n)\) or:

\[
\max_{e,n} \Pr(e + \varepsilon > 0)n\alpha_{FC} - c(e, n), \tag{FC}
\]

where \(\Pr(e + \varepsilon > 0)n\) equals the expected number of accepted articles.

Let \((e^*_{FC}, n^*_{FC})\) be the unique maximizer of the flat-contract optimization problem. We assume the following relationship between the parameters of the two contracts:

\[
\alpha_{FC} = \beta_{LC} \mathbb{E}[e^*_{FC} + \varepsilon|e^*_{FC} + \varepsilon > 0]. \tag{1}
\]

In other words, the expected payment per published article under the LC contract is equal to the per-article flat-rate payment. This follows from the way the per-pageview payout for the linear contract (i.e. \(\beta_{LC}\)) was determined by our partner firm. The firm chose \(\beta_{LC}\) so that the total expected per-article payout under the linear contract if calculated using the average per-article pageviews under the flat contract would equal the flat-contract per-article payment \(\alpha_{FC}\).

Now consider the linear contract. Since an unapproved article receives zero pageviews, the expected number of pageviews conditional on \((e, n)\) equals

\[
\mathbb{E}[\max\{0, e + \varepsilon\}] = \Pr(e + \varepsilon < 0) \cdot 0 + \Pr(e + \varepsilon > 0)\mathbb{E}[e + \varepsilon|e + \varepsilon > 0]
= \Pr(e + \varepsilon > 0)\mathbb{E}[e + \varepsilon|e + \varepsilon > 0].
\]

\(^{14}\)For now, we are assuming that the marginal product of effort is fixed and known, and the only way for writers to increase the expected number of pageviews is to increase their effort. We will add an experimentation aspect to the framework, where certain possible action have ex-ante unknown marginal productivity of effort.

\(^{15}\)The actual pay-for-performance contract implemented by the firm is piece-wise linear in the number of users and is described in detail in the following section. For now we assume a linear contract with constant slope \(\beta_{LC} > 0\) for ease of exposition.
Under the linear contract then, the writer solves the following:

\[
\max_{e,n} \text{Pr}(e + \varepsilon > 0) \mathbb{E}[e + \varepsilon|e + \varepsilon > 0] n\beta_{LC} - c(e, n)
\]  

Proposition 1. The writer is at least as well off under the linear contract than she is under the flat contract.\(^{16}\)

Proposition 2. Assume that \(\varepsilon\) is distributed uniformly on \([-x, x]\), where \(x > 0\) is chosen to be large enough to guarantee that for all possible optimal \(e^*\), we have \(\text{Pr}(e + \varepsilon < 0) > 0\). Assume also that \(\frac{\partial^2 c(e, n)}{\partial e \partial n} > \frac{\alpha_{FC}}{2x}\). If \((e_{FC}^*, n_{FC}^*)\) and \((e_{LC}^*, n_{LC}^*)\) are the choice variables that are the solutions of each of the two optimization problems under these assumptions on \(\varepsilon\), then:

\[
e_{LC}^* \geq e_{FC}^* \text{ and } n_{LC}^* \leq n_{FC}^*.
\]

Several testable implications emerge from the model: first, average per-article effort increases under the linear contract. Second, the number of submissions decreases under the linear contract. Further, article-level quality should increase on average as a result of increased effort. The model can be extended to incorporate multiple actions with differential marginal productivities of effort (both known and based on writers’ beliefs).

### 5 Data

This section describes the data that we use to analyze our various research questions. We draw from several data sources: the main source is the firm’s administrative data (described in Section 5.1), but we also have access to the responses to a battery of questions that writers must complete when they submit an article, and another set of questions that editors must complete when they review an article (described in Section 5.2). We further generate additional data using readability score algorithms and crowd-sourced ratings and evaluations of the articles (still to be done, but an outline can be found in Section 5.3). We then describe our primary outcome variables in Section 5.4.

\(^{16}\)Proofs of Propositions 1 and 2 are presented in Appendix A.
5.1 Observed measures

The digital submission framework (workbench) and website data allow us to match page-level measures from Google Analytics (such as pageviews and users) with individual article-level and writer-level information. The website’s log files for writer profiles, articles, and payments are indexed and searchable using ElasticSearch, which we use to create writer- and article-level measures of input decisions and article impact. User and pageview values are matched with Google Analytics using page titles. While we employ users and views interchangeably, note that the actual definitions differ: “users” measures the number of people (IP addresses) that have viewed a page (article) at least once in a day; pageviews measures the number of views on a page by session. For a single page, these numbers differ whenever a user has viewed the same page more than once and only if that is done in more than one session (sessions restart after thirty minutes of inactivity). The two measures are highly correlated ($\rho = .987$).

Having matched each page with article-level information, we can obtain each article’s publication status (published, rejected), date posted and published, title, text, revision history, category (politics, news, etc), county (the firm allows geographical tags for a number of counties), and the writer’s ID. We subsequently match writer IDs with the writers’ profile information (including treatment status). Combining these data sources allows us to use writers’ submission and publication histories to create writer-level measures including average impact, tenure, and weeks active.

The firm’s database and Google Analytics data begin in early 2014. For the purpose of this study, all writer and output data are collected beginning in September of 2016, with the exception of writer tenure at the firm. As discussed in Section 3.2, all 480 writers that published at least one article since September 5, 2016 were included in the randomization. Our dataset represents all articles and output data for the firm except for editors.\footnote{Editors that also publish articles were excluded from the study because they submit primarily live news feeds (such as election updates) and have a privileged position in the publication process. On average, editor articles perform higher than the writer average but are published much less frequently.}

5.2 Self-Reported Measures

In addition to the administrative data, we collect several self-reported measures from the writers each time they submit an article. Writers are prompted upon submission to (i) approximate the
number of minutes that they spent putting together the news story and (ii) predict how many pageviews they think the article will get in its first week. Writers are assured in writing that editors cannot view their answers to these questions and that the responses will not affect an article’s publication chances. Rather, they are encouraged to report truthfully so that “the firm can learn to better serve them.”

Writers are able to access their profile pages, which provide up-to-date pageview counts for each published article, and can be accessed by the writer and readers by clicking on the profile link in an article’s byline. Given that writers can access article-level pageviews and that users and pageviews are highly correlated, we argue that writers’ predicted pageviews constitute a reasonable measure of writers’ expectations about the profitability of an article.\(^\text{18}\) Collection of beliefs and effort data began on July 27, 2017. We therefore have a little over a month of data on expectations prior to the introduction of the treatment (for the 127 writers that submitted during this period).

We also asked editors a set of questions about articles that they were reviewing. First, they were asked to rate the article on a scale from 1-5, with one representing the highest quality and 1 the lowest. Second, editors had to predict how many pageviews they think the reviewed article will get in its first week. We believe that editors took this task reasonably seriously, and will use the first measure as our first measure of article quality.

### 5.3 Generated data

#### 5.3.1 Readability scores

To explore the effects of the contract change on content, we construct several measures: first, we use the article text to construct five “readability” measures that allow us to examine if writers change the something about how they write following the introduction of the PPV contract. “Readability” generally refers to the level of vocabulary and sentence structure (Chall and Dale, 1995; DuBay, 2004), and these measures employ various approaches to evaluate this. Most measures use some function of word frequency and sentence length to construct a text-level score. To measure read-

\(^{18}\)We did not include financial incentives to encourage accurate time and belief responses. Time is not verifiable and there is mixed evidence that incentivized belief elicitation results in accurate reporting. In a recent study, Hoffman and Burks (2017) compare predicted productivity for truck drivers (miles per week) using both incentivized and unincentivized elicitation. They found no evidence that reported beliefs were different when drivers are rewarded for accuracy.
ability, we focus on the number of syllables, the number of words, and three common and reliable scores in the literature: Flesch Reading Ease, Coleman-Liau, and Dale-Chall.

The Flesch Reading Ease formula is measured on a scale from 0 to 100, with zero being the most difficult to read. The other two measures use textual characteristics to generate grade levels based on an estimate of the number of school years needed to comprehend the text. We multiply these values by negative one so that higher values indicate clearer writing across scores. The Flesch Reading Ease is based on syllable counts while the Coleman-Liau index uses character counts (letters) relative to the number of sentences. The Dale-Chall formula compares the text to a list of 3,000 “easy” words. The formulas used in to calculate the scores vary in their constants and the components used to rank the readability, but generally provide similar rankings across texts.  

5.3.2 Bias and quality ratings

Concerns about the implications of performance pay in the media context often center on the potential ramifications for media polarization and political accountability. We are therefore keen to explore the effect that these new contracts have on journalistic bias. In analyzing this effect of the intervention will also allow us to explore the relative strength of demand-side vs. supply-side sources of media bias. Theories of supply-side sources of media bias suggest a role for wages and journalistic discretion on the incentives for reporters to bias news to promote their world view (Baron, 2006), while demand-side theories argue that bias only occurs in equilibrium if firms find it profitable to distort information to match consumers’ ideologies (Gentzkow and Shapiro, 2006; Mullainathan et al., 2005). Since the control group has little incentive to respond to reader preferences, we assume that any increase in bias or polarization in the treatment group would be due to demand-side factors, i.e. the political preferences of consumers.

To measure this bias, we will enlist auditors on MTurk to classify the articles into politics or not-politics. Auditors will then be asked to identify the political parties mentioned in the article. For each party that is mentioned, auditors will answer the question "Is this article generally positive, neutral, or negative towards this political party?" using a 5-point scale. Answers will then be averaged across auditors (three per article) and normalized to generate party-favorability scores.

---

See Hengel (2016) for further discussion on the comparability of the scores and the criticisms of their accuracy by context.
for each article labeled politics. Using this training set, we will use machine learning modules to generate measures of political bias for the whole corpus of articles.

However, since writers choose what topics to write about, a writer who chooses to write about topics that are consistently negative about a specific ethnicity/party is likely biased. Since negative news or events are time-specific, we will define writers as biased if their articles are consistently more negative about a party than the average article that week. Specifically, we will define a party-bias dummy variable, \( party_{bias_{it}} \), which equals 1 if writer \( i \)'s average article is more negative about that party than the average article that week. Using this approach, we will be able to assess writer slant, which in this case encompasses both the issues that writers choose to cover, and whether their coverage is perceived as fair or slanted.

5.4 Outcomes

We estimate treatment effects using writer-months as the primary level of observation. We had originally planned (and pre-specified) to use writer-weeks, but because writers drop in and out of being active between weeks, thereby changing the sample composition week to week, we felt that the interpretation of results was more difficult at the week-level.

Writer-level measures of monthly outcomes post-treatment are constructed using within-writer values at the month level. For example, the number of published articles for writer \( i \) at time \( t_1 \) is the count of all published articles for writer \( i \) in the first week post-treatment (\( t = 1 \)). We begin by thinking about the implications of the new contracts on the firm’s objective function by exploring the total number of views that treatment writers obtain compared to control writers. We also examine views per article and the variable cost per click. The latter is of interest since the firm’s revenues are directly proportional to pageviews. We then turn our attention to the effects of the new contracts on content: the types and quality of articles published. We have four measures of this for the time being: politics vs. not-politics, local news vs. national, editor-rated quality, and a click-bait score. We will supplement this with the bias measurements from the MTurk/machine learning exercise. Finally, we examine writer behavior by looking at the number of articles submitted and published, as well as article-level and monthly effort. These two are direct tests of Proposition 2 from our model, i.e. we interpret writers’ predicted pageviews as the mean of their subjective beliefs distribution conditional on article qualities (such as topic and time of submission).
In the next version of the paper, we will also examine selection, writer welfare, and heterogeneity but that is yet to come!

6 Results

6.1 Graphical results

We begin our results discussion by showing graphically some of the key outcome variables, by treatment group. Figure 4 compares the total article views in the control and treatment groups before and after the treatment. The first horizontal line depicts the beginning of the treatment, and the second line depicts the last week of the treatment. For the first half of the treatment period, it is clear that the PPV contract resulted in an immediate, sharp jump in the total number of pageviews for the treated writers. These impacts do appear to diminish towards the second half of the period.

![Figure 4: Total views by treatment - weekly](image)

Total pageviews are of course a combination of the number of articles published and the average views per article. Figures 5 and 6 therefore look at the average views per article by treatment group and the number of articles submitted. From these two graphs, it appears that the PPV incentives are resulting in high-impact articles on average, with immediate and persistent effects. However, the number of articles submitted in the PPV group declines over the second half of the treatment.
period both relative to pre-treatment levels and relative to the control group.

Figure 7 shows the variable cost per click in the two groups. Since firm revenues are directly proportional to views, this is a metric of interest. The control group continue to be paid the fixed piece-rate over the time period, so when overall views decline in the later half, these articles become extremely expensive. One interpretation is that the PPV contracts acts as an anti-cyclical protection for the firm, while the piece-rate insures the writers against fluctuations in overall readership.

![Figure 5: Average users per article, by treatment – weekly](image1)

![Figure 6: Average submitted articles by treatment - weekly](image2)

We can also examine some of the questions about the impact of performance pay on content
Figure 7: Cost per click, by treatment – weekly

graphically. Figure 8 depicts the average editor ratings between the two contracts, and suggests that editors generally seem to evaluate the PPV articles to be of higher quality. Whether or not a reader finds this result surprising likely reflect their priors about what news consumers want, and how journalists are likely to respond to click-based contracts more generally.

![Average editor-rated quality by treatment - weekly](image)

Figure 8: Average editor-rated quality by treatment - weekly

We can also examine the clickbait score that articles receive, as measured by an algorithm that is admittedly imperfect for the setting (trained on outlets like the New York Times vs. Buzzfeed; we are in the process of re-training it on Kenyan equivalents). Figure 9 shows these results. Strikingly,
the two groups appear very similar on this metric, but we don’t know whether this is measurement error due to an improperly trained algorithm.

Figure 9: Average clickbait score by treatment - weekly

The next section tests the theoretical hypotheses from Section 4 econometrically, allowing us to control for a variety of controls, writer fixed effects, etc. We will also explore whether the impacts are driven by the selection of writers that contribute following treatment, or by behavioral changes made by writers.

6.2 Average treatment effects

We begin our analysis by estimating the average impact of assignment to the PPV contract on writer submissions and article impact. Relative to a piece-rate contract, rewarding outputs may be more efficient when agents have specific knowledge of the production function and when these benefits outweigh the cost of compensating agents for increased risk. In the absence of learning, treatment writers should shift their effort or article characteristics to increase their revenues. We test whether the PPV contract changes writers’ behavior along a number of dimensions and report both the intent-to-treat (ITT) estimates (i.e. simply comparing the treatment effects across groups) and the effects on active writers.

Following McKenzie (2012), we pool the data and estimate the following ANCOVA model to estimate the average treatment effects, controlling for baseline outcomes $Y_{i,pre}$, individual char-
acteristics $X_i$, strata fixed effects $\nu_s$, and month fixed effects $\delta_t$ for $t = 1, 2, \ldots, T$ post-treatment months:\(^{20}\)

\[
Y_{it} = \sum_t \delta_t + \gamma_1 PPV_{it} + \gamma_2 Choice_{it} + \phi \bar{Y}_{i,\text{pre}} + X_i + \nu_s + \epsilon_{it} \tag{2}
\]

where $Y_{it}$ is constructed at the writer-month level and the sample is balanced for the post-treatment weeks (i.e., values are set to zero in months during which the writer did not submit). All the results are highly robust to conducting the analysis at the week-level, but we chose months instead to reduce the extent to which we are comparing different samples in the different time periods. The pre-treatment outcomes ($\bar{Y}_{i,\text{pre}}$) are calculated by summing the outcome variable for writer $i$ over eight weeks prior to the treatment and dividing by eight to create a within-writer weekly measure.\(^{21}\)

For example, if a writer submitted at least one article in two of the eight weeks prior to treatment, their baseline value of the likelihood of submitting per week would be 0.2. We estimate average treatment effects (ATEs) for the following outcome variables at the month level: the total number of views, the average views per article published, the average cost per click, the number of articles submitted, number of article published, effort (both per article and per month), quality, clickbait scores, and topic choices (politics vs. non-politics). We cluster our standard errors at the writer-level (the level of randomization) in all estimations.

Table 2 reports the ITT estimates for the full sample of 480 writers over the post-treatment months. All outcomes are reported at the writer-month level. Overall, total views in the PPV treatment group are roughly double those in the control group, even though we assign zeros to those who publish nothing. Views per article are close to 150 percent greater in the PPV group than in the control group, and the cost per click falls drastically in the treatment group.

Turning to Table 3, we can see that writers in the PPV treatment submitted almost 27 fewer articles per month, and published roughly 23 articles less than control writers (compared to the control mean of 50 and 42, respectively). Our model suggests that effort per article will increase under the PPV contract conditional on the production and cost functions specified (Proposition 2).

\(^{20}\)Alternatively, we run difference-in-differences specifications presented in Appendix B.

\(^{21}\)We choose to use eight weeks of pre-treatment data due to the the degree of autocorrelation in the outcome variables. Writers that are not active in those eight weeks are assigned values of zero for the baseline means and we include a dummy variable to control for non-activity. We report estimates using eight weeks of pre-treatment data but the results are not sensitive to the time frame chosen.
We test whether total effort and effort per article indeed increase in response to the PPV contract using the self-reported measure of the number of minutes spent producing each article. While effort per article appears to be unaffected, we do see a sharp drop in monthly effort in column (4). The results from Tables 2 and 3 combine to provide evidence that the treatment contract increased total and per-article impact \((q_{LC} > q_{FC})\), as predicted by the model. The substantial difference in cost per click suggests that the PPV contract had a positive impact on firm revenue. Further, PPV writers produced higher-impact articles, precisely as the firm had hoped when designing the contract. Clearly, these estimates are partial-equilibrium effects and only provide suggestive evidence of the impact of the treatment on firm revenues and profits. Currently, we are unable to determine how much of the increase in productivity in the PPV treatment was due to substitution of users away from articles in the control group or due to attracting readers from other sources.

We also examine treatment impacts on content. Table 4 reports these results. Contrary to

\(^{22}\)The effort regressions (columns (3) and (4) in Table ?? present results where self-reported minutes have been winsorized at the 1st and 99th percentiles to reduce the influence of observations with inordinately high minutes due to the self-reported nature of the data. The affected outliers report over 10,000 hours per article. We report estimations including outliers in the Appendix.
Table 3: ITT estimates of treatment on writer submissions and effort

<table>
<thead>
<tr>
<th>Variables</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N Submitted</td>
<td>N Published</td>
<td>Minutes/article</td>
<td>Minutes/month</td>
</tr>
<tr>
<td>PPV</td>
<td>-26.65***</td>
<td>-22.75***</td>
<td>4.440</td>
<td>-457.9*</td>
</tr>
<tr>
<td></td>
<td>(7.152)</td>
<td>(6.271)</td>
<td>(3.332)</td>
<td>(243.8)</td>
</tr>
<tr>
<td>Tenure</td>
<td>0.313**</td>
<td>0.289*</td>
<td>-0.0437</td>
<td>7.653**</td>
</tr>
<tr>
<td></td>
<td>(0.153)</td>
<td>(0.153)</td>
<td>(0.0608)</td>
<td>(3.756)</td>
</tr>
<tr>
<td>Baseline</td>
<td>0.657***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0814)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td></td>
<td>0.627***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0714)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td></td>
<td></td>
<td>0.509***</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.0608)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>22.91*</td>
<td>15.95</td>
<td>16.31**</td>
<td>794.7**</td>
</tr>
<tr>
<td></td>
<td>(13.43)</td>
<td>(11.95)</td>
<td>(6.489)</td>
<td>(325.1)</td>
</tr>
<tr>
<td>Observations</td>
<td>317</td>
<td>317</td>
<td>280</td>
<td>280</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.460</td>
<td>0.458</td>
<td>0.481</td>
<td>0.646</td>
</tr>
<tr>
<td>Month FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Strata FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Mean Dep. Control</td>
<td>49.29</td>
<td>42.38</td>
<td>32.05</td>
<td>1351</td>
</tr>
</tbody>
</table>

Standard errors clustered at the individual level
Sample: Writer-months following treatment (Unbalanced)

In the graphical analysis, the impacts of the PPV contract on editor-rated article quality appear to be small and they are not significantly different from zero. Similarly, the impacts on click-bait are statistically indistinguishable from zero. This may again be due to measurement error, but it certainly does not provide strong support for some of the more pessimistic predictions that some media analysts believe will automatically take place when journalists get rewarded based on views.

We do see a sharp uptick in the share of political articles in the treatment group, which may be related to the timing of the treatment period around an intense political time in Kenya. Political articles make up the majority of all submitted articles (58 percent) for this period, and they also tend to receive higher average views than all other categories both before and after the treatment.

This shift in topical focus might reflect that PPV writers either have or are gaining knowledge of the average returns to political articles relative to other categories. We include month fixed effects to rule out impacts driven by increased political activity during the time surrounding the election period. Conditional on a fixed number of political stories in a given week, PPV writers
have a higher incentive to submit high-impact political stories faster than control group writers. Writers in the piece-rate contract do not know about the existence of the PPV treatment group, so it seems unlikely that this effect reflects a substitution away from political articles in the control group. However, it could be that increased competition to submit timely political articles could lead control writers to decrease the share of political articles they submit.

Table 4: ITT estimates of treatment on news content

<table>
<thead>
<tr>
<th>Variables</th>
<th>(1) Avg. Quality</th>
<th>(2) Avg. Clickbait</th>
<th>(3) Share Political</th>
<th>(4) Share County</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPV</td>
<td>0.149</td>
<td>-2.096</td>
<td>0.257***</td>
<td>-0.0964*</td>
</tr>
<tr>
<td></td>
<td>(0.191)</td>
<td>(1.692)</td>
<td>(0.0504)</td>
<td>(0.0564)</td>
</tr>
<tr>
<td>Tenure</td>
<td>-0.00512**</td>
<td>-0.0509</td>
<td>-0.000860</td>
<td>0.00214*</td>
</tr>
<tr>
<td></td>
<td>(0.00247)</td>
<td>(0.0324)</td>
<td>(0.00102)</td>
<td>(0.00110)</td>
</tr>
<tr>
<td>Baseline</td>
<td>0.218</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.141)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td></td>
<td>0.179*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0924)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td></td>
<td></td>
<td>0.615***</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.116)</td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td></td>
<td></td>
<td></td>
<td>0.697***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0836)</td>
</tr>
<tr>
<td>Constant</td>
<td>2.800***</td>
<td>30.19***</td>
<td>0.173**</td>
<td>-0.00638</td>
</tr>
<tr>
<td></td>
<td>(0.543)</td>
<td>(3.504)</td>
<td>(0.0860)</td>
<td>(0.0866)</td>
</tr>
<tr>
<td>Observations</td>
<td>264</td>
<td>278</td>
<td>280</td>
<td>280</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.274</td>
<td>0.168</td>
<td>0.455</td>
<td>0.453</td>
</tr>
<tr>
<td>Month FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Strata FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Mean Dep. Control</td>
<td>2.814</td>
<td>30.06</td>
<td>0.334</td>
<td>0.597</td>
</tr>
</tbody>
</table>

Standard errors clustered at the individual level
Sample: Writer-months following treatment (Unbalanced)

We further estimate the impact on total users, users per article, and number of articles submitted, allowing the effect to vary by week. Figures 10-12 plot the resulting coefficients. These graphs overall seem to corroborate the graphical results in Section 6.1.

6.3 Mechanisms

THIS WHOLE SECTION NEEDS UPDATING & REVISING

The previous section analyzed the ATE of the program along a variety of dimensions; this
Figure 10: Treatment effects on total views, by week

Figure 11: Treatment effects on average views, by week
section delves into issues of selection, learning, and experimentation. We begin in Section 6.3.1 by exploring whether the impacts of the PPV contract are due to selection of more productive writers into treatment or the margins over which writers change their behavior in response to the PPV contract. Writers in the PPV treatment decreased the number of articles submitted per week as predicted by the model. In addition, writers’ action set includes content choice, readability, and time spent per article, all of which may interact with one another. Furthermore, writers are presumably learning about the effect that these various strategies have on their payoffs. To explore this aspect of the writers’ production, Section 6.3.2 examines the extent to which writers experiment with different strategies over time. Section 6.4 examines another aspect of learning: writers ability to learn about the production function of pageviews. We study this by looking at writers’ expectations about pageviews, whether there are observable characteristics that they believe affect pageviews, and whether they get better at predicting their pageviews over time.

6.3.1 Selection

In the case of the PPV contract, the theory in Section 4 argues that the PPV contract will induce workers both to submitting fewer articles, and by changing their behavior (effort, experimentation) for the articles that they do write. However, the PPV contract could also affect the selection of writers who choose to leave the firm. e.g., output-based pay may deter writers with low productivity
or a distaste for pay variation from submitting at all. On the other hand, the new contract may attract writers that did not plan to submit again under the per-article contract. Those new writers will not appear in our sample, however, but we can analyze them separately.

Our first test is to examine whether writers in the PPV group dropped out at different rates than the control group following the introduction of the new contract. Of the eighty-six writers who submitted an article in the two weeks prior to the treatment, we explore how many of them submit an article after treatment.

We then analyze the results from a regression of $Submit > 0$ after treatment reveals no statistical impact of the treatment on separation.

**Needs updating.**

Our second test of selection is to explore whether the writers that submitted articles after the treatment appear different on observable characteristics, including pre-treatment impact, beliefs, and tenure. We first compare the total numbers of active writers across treatment groups. We also examine differences between active writers along observable characteristics. For example, if the PPV contract caused selection of out the contract based on average productivity, we would expect baseline users per article to be higher for writers in the PPV group.

**Needs updating.**

Finally, we test for selection using a method similar to Lazear (2000) who includes worker fixed effects in a regression of daily productivity before and after a contract regime change. The Lazear study finds that including fixed effects reduces the estimate of the contract change by more than half; the author argues that the resulting estimate is the incentive effect. In the same vein, Table XX includes writer fixed effects in a difference-in-differences regression of submissions, publications, and views per article.

**6.3.2 Experimentation**

We first test whether the contract type determines the type of content that workers choose to produce over time. We examine whether writers experiment with different topics, article lengths, or by writing about different geographic areas.
In contrast to a piece-rate contract, rewarding writers proportionally to their output may strengthen their incentive to exploit their contextual knowledge, use their existing inputs more efficiently, or innovate through experimentation (if the expected benefits from experimentation exceed the increased risk). The evidence from Sections 6.1 and 6.2 demonstrates that writers adjusted their behavior immediately following the introduction of the treatment. The immediate adjustment suggests that writers’ already had some knowledge based on past experience about the returns to various inputs, such as effort, topic choice, and location choice.

6.4 Predicting pageviews

As described in Section 5.2, we elicit writers’ beliefs about the productivity of each article at the time of submission. To investigate the role of contract type on writers’ ability to predict their payoffs, we first test whether writers in the PPV treatment group are better at predicting productivity after the treatment launch by estimating the following regression, using measures of writers’ beliefs about expected clicks per article prior to submission:

$$\text{Accuracy}_{it} = \delta_t + \gamma_1 \text{PPV}_{it} + \gamma_2 \text{Choice}_{it} + \phi \bar{\text{Accuracy}}_{i,\text{pre}} + \nu_s + \epsilon_{it}$$

where $\text{Accuracy}_{it} = |\hat{v}_{it} - v_{it}|$ is writers’ $i$’s average prediction accuracy in week $t$, $\hat{v}_{it}$ is the average subjective belief about articles in week $t$ stated when submitting the article, $v_{it}$ represents the average views received by articles published in week $t$, $\bar{\text{Accuracy}}_{i,\text{pre}}$ is the baseline average views per article to date, and $X_i$ is a vector of individual controls including tenure.

Estimates of the treatment effects on writers’ predicted pageviews and accuracy are presented in Table XX.
7 Robustness checks

We explore the robustness of our findings in several ways. The first is to examine the possibility that results are entirely driven selection and/or cheating by writers. As it turns out, the checks built in to deal with writers signing up for new accounts were not bullet-proof and we have anecdotal evidence that some writers opened new accounts and were randomized into the control group. Whether this was purposeful or not, we can’t know for sure, but we want to make sure that the results are not driven by this potential “cheating”.

It should not be a concern for the regressions where we examine month-level totals (total pageviews, for example) since we already include writers who don’t publish as zeros. In some regressions, however, we only observe those who publish (pageviews/article, for example), so we follow a procedure outlined in Bazzi et al. and explore what would happen if all the writers who dropped out post-treatment (became inactive) were in the 5th and 95th percentile of the article-impact distribution. We do this by imputing values for those writers, as well as one version where we assume that drop-outs continued to publish at their baseline levels or something.

Needs updating.

8 Conclusions

Using random variation in assignment to contract type for journalists in an online news firm in Kenya, we show that output-based incentive contracts sharply increased average and per-article impact percent, respectively, alongside a substantial reduction in the frequency of article submission. The impacts on productivity and submission intensity are evident and of similar magnitudes using both ANCOVA and difference-in-differences estimations. We still need to explore the nature of selection in this context, but we do observe within-writer behavioral changes and have the ability to further pursue the question of selection.

While we find that the performance contract increased both total and average article-impact, we are not yet able to determine whether the change will lead to an increase in long-term profits, although the much higher cost-per-click in the control group suggests that PPV is advantageous for the firm. In fact, the firm has since the experiment moved all writers to a revised version of
PPV. While the firm’s goal of producing more “high-impact” articles seems to be satisfied, the key unknown in our analysis is the relationship between total content and views is unknown, so we cannot speak to what would happen if the number of articles drops “too far”. Furthermore, the increase in views per article in the performance treatment could be the result of substitution of views away from the control group (competition). In future work, we plan to determine which effect dominates in the short- and long-run.

We have excluded a deeper (forthcoming) discussion on quality and text measures that will be integral in determining how the treatment affects the bias and polarization of news, and its implications for the firm’s reputation and for political accountability.

In future work, we will focus both on the incidence of (better-measured) click-bait and political bias as outcomes of the performance contracts. In addition to using current click-bait and bias detection models, we will recruit workers on Amazon’s mTurk to evaluate title and text discordance and political bias for a subset of articles. Using this training set, we will use several machine learning modules to generate measures of click-bait and political bias for the whole corpus of articles. We will test the degree to which potential bias is driven by demand-side or supply-side bias.
References


36


Appendix A: Proofs of Propositions 1 and 2

Proof. Evaluating the objective function of (LC) at (FC)'s maximizers \((e_{FC}^*, n_{FC}^*)\) and using (1), it equals

\[ \Pr(e_{FC}^* + \varepsilon > 0) n \alpha_{FC} - c(e_{FC}^*, n_{FC}^*) , \]

which equals the maximum value of (FC)'s objective function. Thus, under the linear contract, the agent can guarantee herself utility of at least as much as what she can get under the flat contract. 

\[ \square \]

Proof. Given the assumptions on \(\varepsilon\), the optimization problem (FC) becomes:

\[ \max_{e,n} \frac{x + e}{2x} n \alpha_{FC} - c(e, n), \]

while (LC) is

\[ \max_{e,n} \frac{(x + e)^2}{4x} n \beta_{LC} - c(e, n). \]

Finally, (1) implies

\[ \alpha_{FC} = \beta_{LC} \frac{x + e_{FC}^*}{2}. \tag{4} \]

Define the function \(F : \mathbb{N} \times \mathbb{R}_+ \times \{0, 1\} \to \mathbb{R}\) by

\[ F(e, n; \mu) = \begin{cases} \frac{x + e}{2x} n \alpha_{FC} - c(e, n) & \text{if } \mu = 0; \\ \frac{(x + e)^2}{4x} n \beta_{LC} - c(e, n) & \text{if } \mu = 1. \end{cases} \]

The function \(F\) is infinitely differentiable for each value of \(\mu\) and

\[ \frac{\partial^2 F(e, n; 0)}{\partial e \partial (-n)} = \frac{\partial^2 c(e, n)}{\partial e \partial n} - \frac{\alpha_{FC}}{2x}, \]

which is strictly positive by assumption. Therefore \(F\) is quasi-supermodular in \((e, -n)\) for \(\mu = 0\).
By the definition of quasi-supermodularity:

\[ F(e_{FC}^*, n_{FC}^*, 0) \geq F(\min\{e_{FC}^*, e_{LC}^*\}, \max\{n_{FC}^*, n_{LC}^*\}, 0) \]

\[ \Rightarrow F(\max\{e_{FC}^*, e_{LC}^*\}, \min\{n_{FC}^*, n_{LC}^*\}, 0) \geq F(e_{LC}^*, n_{LC}^*, 0) \]  

(5)

where the first inequality follows from the fact that \((e_{FC}^*, n_{FC}^*)\) solves (FC).

Let \(\bar{e} := \max\{e_{FC}^*, e_{LC}^*\}\) and \(\bar{n} := \min\{n_{FC}^*, n_{LC}^*\}\). Using the definition of \(F\), we can re-write (5):

\[ \alpha_{FC} \frac{x + \bar{e}}{2x} - c(\bar{e}, \bar{n}) \geq \alpha_{FC} n_{LC}^* \frac{x + e_{LC}^*}{2x} - c(e_{LC}^*, n_{LC}^*) \]

\[ \Rightarrow \alpha_{FC} \frac{x + \bar{e}}{2x} - \alpha_{FC} n_{LC}^* \frac{x + e_{LC}^*}{2x} \geq c(\bar{e}, \bar{n}) - c(e_{LC}^*, n_{LC}^*) \]  

(6)

Clearly, \(\bar{e} \geq e_{FC}^*\), and \(\bar{n} \geq e_{LC}^*\). If \(e_{FC}^* \geq e_{LC}^*\), using (4) we have:

\[ \beta_{LC} \frac{(x + \bar{e})^2}{4x} - \beta_{LC} n_{LC}^* \frac{(x + e_{LC}^*)^2}{4x} \]

\[ \geq \alpha_{FC} \frac{x + \bar{e}}{2x} - \alpha_{FC} n_{LC}^* \frac{x + e_{LC}^*}{2x} \]

\[ \geq c(\bar{e}, \bar{n}) - c(e_{LC}^*, n_{LC}^*) \]

where the last inequality follows from (6).

If \(e_{FC}^* < e_{LC}^*\) instead, then:

\[ \beta_{LC} \frac{(x + \bar{e})^2}{4x} - \beta_{LC} n_{LC}^* \frac{(x + e_{LC}^*)^2}{4x} \]

\[ = \beta_{LC} \frac{x + e_{LC}^*}{2} \left( \bar{n} \frac{x + \bar{e}}{2x} - n_{LC}^* \frac{x + e_{LC}^*}{2x} \right) + (\bar{e} - e_{LC}^*) \left( \beta_{LC} \frac{x + \bar{e}}{4x} \right) \]

\[ \geq c(\bar{e}, \bar{n}) - c(e_{LC}^*, n_{LC}^*) \]  

39
Either way, this is equivalent to:

\[ F(\bar{\epsilon}, n, 1) \geq F(\epsilon_{LC}^*, n_{LC}^*, 1). \]

But since \((\epsilon_{LC}^*, n_{LC}^*)\) is the unique maximizer of \(F\) for \(\mu = 1\), this implies:

\[ \bar{\epsilon} = \max\{\epsilon_{FC}^*, \epsilon_{LC}^*\} = \epsilon_{LC}^* \] and \(n = \min\{n_{FC}^*, n_{LC}^*\} = n_{LC}^*; \]

which is what we wanted to show. \(\square\)
Appendix B

Needs updating.