RUHR
ECONOMIC PAPERS

Nadja Kairies
Miriam Krieger

How do Non-Monetary Performance Incentives for Physicians Affect the Quality of Medical Care?

A Laboratory Experiment

UNIVERSITÄT DUISBURG ESSEN

#414

Imprint

Ruhr Economic Papers

Published by

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI)
Hohenzollernstr. 1-3, 45128 Essen, Germany

Editors

Editorial Office

Ruhr Economic Papers #414

Responsible Editor: Volker Clausen

ISSN 1864-4872 (online) – ISBN 978-3-86788-469-3

Ruhr Economic Papers #414

Nadja Kairies and Miriam Krieger

How do Non-Monetary Performance Incentives for Physicians Affect the Quality of Medical Care?

A Laboratory Experiment

UNIVERSITÄT DUISBURG ESSEN

Bibliografische Informationen der Deutschen Nationalbibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über: *http://dnb.d-nb.de* abrufbar.

http://dx.doi.org/10.4419/86788469
ISSN 1864-4872 (online)
ISBN 978-3-86788-469-3

Nadja Kairies and Miriam Krieger[1]

How do Non-Monetary Performance Incentives for Physicians Affect the Quality of Medical Care? – A Laboratory Experiment

Abstract

In recent years, several countries have introduced non-monetary performance incentives for health care providers to improve the quality of medical care. Evidence on the effect of non-monetary feedback incentives, predominantly in the form of public quality reporting, on the quality of medical care is, however, ambiguous. This is often because empirical research to date has not succeeded in distinguishing between the effects of monetary and non-monetary incentives, which are usually implemented simultaneously. We use a controlled laboratory experiment to isolate the impact of non-monetary performance incentives: subjects take on the role of physicians and make treatment decisions for patients, receiving feedback on the quality of their treatment. The subjects' decisions result in payments to real patients. By giving either private or public feedback we are able to disentangle the motivational effects of self-esteem and social reputation. Our results reveal that public feedback incentives have a significant and positive effect on the quality of care that is provided. Private feedback, on the other hand, has no impact on treatment quality. These results hold for medical students and for other students.

JEL Classification: I11, C91, L15, I18

Keywords: Laboratory experiment; quality reporting; feedback; treatment quality; performance incentives

April 2013

1 Both CINCH and University of Duisburg-Essen. – We thank Jeannette Brosig-Koch, Stefan Felder, and the participants of the 2013 Conference of the German Association for Health Economics dggö for their valuable comments on this paper. Financial support from CINCH and from the German Research Foundation (DFG) is gratefully acknowledged. – All correspondence to Nadja Kairies, University of Duisburg-Essen, CINCH and Faculty of Economics and Business Administration, Universitätsstr. 12, 45117 Essen, Germany, E-Mail: Nadja.Kairies@ibes.uni-due.de.

1. Introduction

Recent healthcare reforms in various countries have specifically aimed at improving the quality of medical care while simultaneously controlling costs (McCellan, 2011). In this context non-monetary performance incentives, predominantly in the form of public quality reporting, have gained increasing popularity among policy makers as a means to achieving these two seemingly contrary goals (Dranove and Jin, 2010).

Evidence on public quality reporting in medical care shows that while it leads physicians to change their provision behavior (Kolstad, 2013), it is not clear that this actually improves the quality of care (Marshall, 2000). On the one hand, there is some indication that public quality reporting can contribute to decreased mortality rates; see Hannan et al. (1994) or Rosenthal et al. (1997). On the other hand, public reporting can also lead to unintended problems, such as a shift in effort towards those aspects of medical care that are reported on and away from unreported aspects (Werner et al., 2009), or the selection of patients towards those whose treatment improves the reported outcomes (Dranove et al., 2003, Cutler et al., 2004, Werner and Asch, 2005). Another issue is that regional characteristics influence the effects of public quality reporting on the quality of medical care: public quality reports have a larger impact, for instance, the more competitive the health care market is (Grabowski and Town, 2011). A further issue with previous empirical studies on the effects of non-monetary performance incentives in health care is that they are often difficult to disentangle from those of monetary incentives, especially as these two mechanisms are typically implemented together, e.g. in the US Premier Hospital Quality Incentive Demonstration within Medicare and in the UK Quality and Outcomes Framework. Simultaneous implementation of multiple new incentives makes it very difficult to establish what the individual impact of each of these changes to the system is, and whether they are in fact substitutes or complements (Maynard, 2012). Hence, Cutler et al. (2004) point out that more research on such incentives is essential to understanding the underlying mechanisms that drive changes in physician provision behavior.

From a theoretical point of view, it is important to differentiate between the modes of performance incentives, i.e. whether feedback is given privately or in public. Private feedback is a competitive incentive which addresses an individual’s self-esteem. Bénabou and Tirole (2002) state that the mere possibility of receiving positive feedback can motivate an individual to increase his performance. Making someone’s relative performance known to others, however, adds a reputational or image aspect to the incentive and speaks to the individual’s desire to gain social status and avoid social disapproval (see Bénabou and Tirole, 2006). On the other hand, monitoring performance and giving feedback also implies control, which can potentially crowd out pro-social behavior (see Ellingsen and Johannesson, 2008).

So far there is barely any empirical research specific to the health care sector which distinguishes the effect of (private) performance feedback based on self-esteem from that (public) based on social reputation. Hibbard et al. (2003) and Hibbard et al. (2005) report on an experimental field study in which they examine the effects of private as well as public feedback on the quality of care in hospitals. Their design includes two intervention groups, one of which receives both private and public feedback and one only private feedback, and a control group which is given no feedback at all. They find that hospitals which receive public feedback are significantly more involved in quality improvement efforts than hospitals with only private or with no feedback (Hibbard et al., 2003). Hospitals in both treatment groups increase their quality compared to the control group hospitals with no feedback, although the differences in average performance changes between the two treatment groups were not statistically significant (Hibbard et al., 2005). However, the results of these studies are subject to some methodological limitations, such as non-random assignment of hospitals to the groups and reliance on self-reported performance measures.

In non-medical settings, evidence for the positive impact of private feedback on performance is provided by several laboratory experiments, for instance Charness et al. (2011) and Kuhnen and Tymula (2012) for output in real-effort tasks. The positive impact of rank information on performance has also been documented in various field studies, including Mas and Moretti (2009) among factory shift workers, Azmat and Iriberri (2010) for high school students, and Blanes i Vidal and Nossol (2011) for white-collar workers. However, studies by Hannan et al. (2008), Eriksson et al. (2009), and Barankay (2011a and 2011b) all report results from laboratory or field experiments which suggest that feedback affects performance either negatively (at least for some individuals or under some conditions) or not at all. There is also evidence from laboratory experiments for a positive impact of public feedback on performance in non-medical settings, such as contributions to a public good (Rege and Telle, 2004) or donations to charity (Ariely et al., 2009). This effect has also been found in the field: In a study of Vietnamese language students, Tran and Zeckhauser (2012) report that both private and public feedback significantly raise test performance as compared to giving no feedback at all. In fact, students who were given public feedback outperformed those who received private feedback, though the difference is only marginally significant.

The lack of evidence for the health care market and the ambiguous results of performance feedback in other domains obscure the picture of how feedback incentives might work in a medical setting. The relevant studies that do exist in the health domain suffer methodological shortcomings, such as reliance on self-reported measures and non-random assignment to intervention groups (Hibbard et al., 2003, and Hibbard et al., 2005). The contribution of this paper is to disentangle the underlying mechanisms of private and public feedback incentives in the medical context in a controlled laboratory experiment. This method allows us to isolate the impact of feedback on the quality of medical care from other factors in the physician’s decision environment, such as the simultaneous variation of financial incentives, regional system characteristics, and the health status of patients. Moreover, laboratory experiments are an inexpensive method to analyze the effects of a planned reform before it is implemented, and can thus help policy makers avoid costly failures. Specifically, our research adds to the literature discussed above in two main ways: Firstly, we investigate how non-monetary performance incentives for physicians affect the quality of the medical services they provide. Secondly, we control for the different motivation mechanisms behind public and private feedback by implementing the two separately and comparing their respective impact on the quality of medical care provided.

In our experiment subjects take on the role of physicians and make decisions over the medical treatment of patients, receiving feedback on the quality of care they provide. To account for the character of a political reform, we employ a within-subject design: In part 1 of the experiment subjects decide on the quantity of medical treatment they provide for a number of patients and are remunerated based on a fee-for-service schedule. In part 2 subjects are asked to make the same treatment decisions for an equal number of patients with the same characteristics as in part 1, but this time they will receive feedback on their performance at the end of the experiment in addition to the remuneration. Physician performance is measured in terms of outcome quality of care for the patient and is fully observable, i.e. not self-reported. Feedback is given in form of competitive rankings and is either private or public. Subjects who receive private feedback are informed about (only) their position in the ranking of participants on their computer screen. For public feedback subjects are asked to stand up while the ranking is read out loud by the experimenter, a procedure similar to that used in experimental studies by Rege and Telle (2004) and Ariely et al. (2009). In order to account for potential professional effects, we compare the decisions made by medical students – physicians in training – to those of other (student) subjects. Patient benefits realized in the experiment accrue to real patients as they are transferred to an organization which provides eye cataract operations.

In section 2 of this paper we describe our experimental design. In section 3 we present results, while section 4 discusses some policy implications and concludes.

2. Experimental Design

Our experiment consists of two parts, each containing a choice task with 9 decision situations. All subjects hence made a total of 18 individual decisions.

Decision Situations

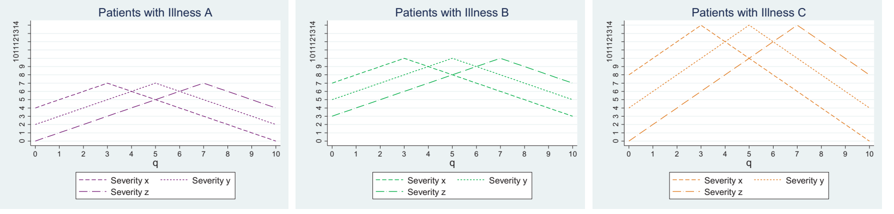
The basic decision situation follows that of Brosig-Koch et al. (2013a, b).[1] The subject takes on the role of a physician and decides on the treatment of a patient. Treatment is performed by allotting the patient a quantity of $q \in [0,1,2, \dots 10]$ medical services. With each treatment decision, the physician simultaneously determines his own profit $\pi(q)$ and the patient's health benefit $B_{kl}(q)$, both measured in monetary terms. For each treatment quantity, the physician also incurs costs $c_{kl}=0.1q^2$ which are deducted from his fee-for-service (FFS) remuneration $R=2q$.[2] This basic decision is repeated sequentially for nine patients, who differ in the benefit they stand to gain from medical treatment.[3] Each patient suffers from one of three illnesses, $k \in [A,B,C]$, which determines the maximum benefit he can receive from optimal treatment ($B_{Al}(q^*)$,= 7, $B_{Bl}(q^*)$,= 10, $B_{Cl}(q^*)$,= 14; see Figure 1). The illnesses each take on one of three degrees of severity, $l \in [x,y,z]$, which in turn determines the quantity of medical services at which a patient gains the optimal benefit from treatment ($q_x^* = 3$, $q_y^* = 5$, $q_z^* = 7$). See Appendix B.1 for a complete set of the parameters adapted from Brosig-Koch et al. (2013a).

[1] Brosig-Koch et al. (2013b) study the effects of pay-for-performance incentives on physicians' provision behavior. Basing our experimental design on theirs allows us to compare financial and non-monetary incentive mechanisms in future research.

[2] We use FFS as it is the principal remuneration structure for primary physicians in most countries, e.g. in the US (Medicare), Australia, France, and Germany. Using a different payment structure such as capitation would presumably not change the qualitative results of our experiment, as we are concerned with a reform which is independent of monetary remuneration.

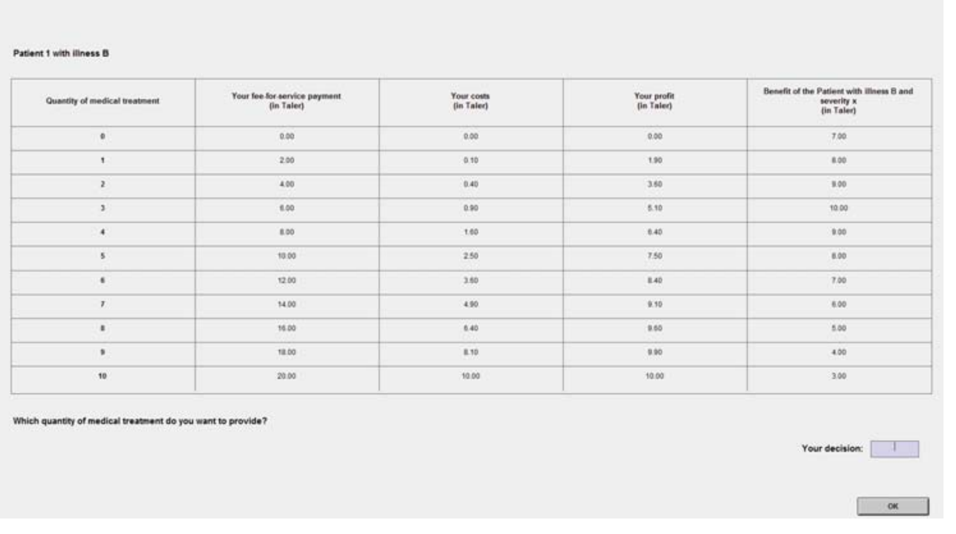
[3] The order of the 9 patients was determined randomly at the outset of the experiment and then kept constant for all subjects and in all variants of the choice task.

Figure 1: Patient benefit functions for illnesses k and severities l



The physician's profit-maximizing choice in every treatment decision is to provide the largest possible quantity of 10 medical services. As this quantity is always higher than the quantity that maximizes the patient's benefit (due to the fee-for-service remuneration scheme), subjects face a trade-off between the two welfare functions in each treatment decision. See Figure 2 for an example of the decision situation.

Figure 2: Example of a decision screen in treatments



Patient 1 with illness B

Quantity of medical treatment	Your fee-for-service payment (in Taler)	Your costs (in Taler)	Your profit (in Taler)	Benefit of the Patient with illness B and severity x (in Taler)
0	0.00	0.00	0.00	7.00
1	2.00	0.10	1.90	8.00
2	4.00	0.40	3.60	9.00
3	6.00	0.90	5.10	10.00
4	8.00	1.60	6.40	9.00
5	10.00	2.50	7.50	8.00
6	12.00	3.60	8.40	7.00
7	14.00	4.90	9.10	6.00
8	16.00	6.40	9.60	5.00
9	18.00	8.10	9.90	4.00
10	20.00	10.00	10.00	3.00

Which quantity of medical treatment do you want to provide?

Your decision:

OK

Patients

The patients in our experiment were not physically present in the laboratory. Nevertheless, the monetary value of the patient benefit went to real patients outside the laboratory. We follow Hennig-Schmidt et al. (2011) and Brosig et al. (2013a, b) in this approach to making patient outcomes in the decision situations directly relevant to health, rather than mere monetary payments. Subjects were instructed that the sum of all patient benefits achieved in the situations selected for payment would be transferred to the charity organization *Christoffel Blindenmission*, which provides care for patients with eye diseases.

Payment

All monetary amounts in the experiment were designated in the experimental currency of Taler. 1 Taler equals € 0.80. In keeping with experimental best practice, one decision situation for each part of the experiment was drawn at random at the end of the experiment (random payment technique) in order to avoid wealth and averaging effects.[4] The situations chosen in each session are valid for all its participants. Each subject received the combined physician profits achieved in these two situations as payoff for the experiment. The benefit received by the patient in these two situations was donated to the *Christoffel Blindenmission*. The donation was carried out immediately after the experimental session was completed and was witnessed by a randomly chosen subject (who received an additional payment of €5 for this task).

Treatment Conditions

In order to address our research questions, we conduct two separate treatment conditions: (1) PRIVATE and (2) PUBLIC.

In condition (1) PRIVATE, the first part of the experiment consists of the choice task as described above: subjects decide on medical treatment for 9 patients. In the second part of the experiment, subjects again make the same treatment decisions for these 9 patients. However, before beginning part two of the experiment, they are informed that at the end of this task, all participants in the session (typically 12 subjects) will be ranked according to the quality of treatment they provide. Treatment quality is defined as the (negative) difference between the realized patient benefit and the optimal patient benefit. The highest treatment quality is thus achieved by choosing the patient-optimal quantity of medical services; in this case treatment quality is zero. This performance feedback is given in private, so that subjects learn only their own position in the ranking (on their computer screen), but not anyone else's. Ranks are shared if participants provide equal treatment quality. Feedback is provided only for the one decision situation in this part of the experiment which has been randomly selected for payment.

Condition (2) PUBLIC is analogous to condition (1), consisting of the basic choice task in part one of the experiment and a feedback incentive for the choice task in part two. Again, subjects are told in the instructions for part two that they will be ranked according to the quality of treatment provided in the situation chosen for payment. In this condition, however, the ranking is made public among the participants of this session: First, the rank table with all participants (identified by their seat numbers) is displayed on their computer screens (see Figure 3). Next, in a procedure similar to that of Ariely et al. (2009) and Rege and Telle (2004), subjects are requested to stand up (allowing everyone to see everyone else over the walls of their cubicles). The ranking is then read aloud by the

[4] Various studies confirm that the random payment technique does not dilute the power of the monetary incentive for non-complex choice tasks (Starmer and Sugden, 1991, Cubitt et al. 1998, Laury, 2006, Baltussen et al., 2010).

experimenter. As they are called up, subjects are required to raise a sign displaying their seat number.

We use a within-subject design to account for the character of a reform that introduces performance feedback. This allows us to analyze behavior before and after the reform in a controlled way: Comparing the decisions made in part one in these two treatments to those made in the incentivized tasks in part two (within-subject comparison) permits us to address our research question Q1 whether feedback incentives have an impact on the quality of medical treatment provided. The comparison of choices made in part two between treatments (1) and (2) helps us answer our research question Q2 whether the mode of delivering feedback – privately or publicly – affects the impact of the feedback incentive on treatment quality.

Figure 3: Example public feedback screen

Medical Students

In all sessions of our experiment, we recruited medical students as well as students of other degree programs as subjects. Comparing decision behavior between these groups allows us to clarify whether prospective physicians – who have perhaps selected themselves into medical education based on specific social preferences, or are influenced by medical professional norms in the course of their training, or both – react differently to reputation-based performance incentives. Ahlert et al. (2012), for example, find that behavior in situations framed as medical treatment decisions (rather than neutral decisions) is impacted by the professional norms of medicine or economics adopted by their subjects. However, other experiments carried out at the Essen Laboratory for Experimental Economics involving different types of health-related decisions have not confirmed this type of professional effect (e.g. Brosig-Koch et al. 2013a, b).

Robustness Check

We test the robustness of our results against the order in which subjects face the incentivized and non-incentivized tasks. Aside from experimental design considerations, private feedback could have motivating or demotivating effects on provision behavior in the second part of the experiment. We reversed the task order in two sessions for treatment condition (1) PRIVATE FEEDBACK: Subjects here completed part 1 with a private feedback incentive and part 2 without a feedback incentive. Note that we could not test for a reverse task order with public feedback as this would imply the loss of subjects' anonymity in part 1 of the experiment, which compromises subsequent decisions in the non-incentivized task in part two of the experiment.

Experimental Procedure

The experiment was carried out at the Essen Laboratory for Experimental Economics (Duisburg-Essen University) in June 2012 using the specialized software z-tree (Fischbacher 2007). 144 subjects were recruited via ORSEE (Greiner, 2004) and participated in a total of 12 sessions of about an hour each.

Subjects were allocated to seats in the laboratory by a random draw. They received separate written instructions at the outset of each part of the experiment and were given several minutes to read the instructions carefully and to ask clarifying questions. At the beginning of part 1, subjects also completed several control questions (see Appendix A) which served to ensure that all subjects understood the task at hand. The control questions were announced in the instructions and were not relevant to any payments earned in later decisions.

At the end of the experiment all subjects were paid out individually and in private. They received an average payoff of €13.51 (min: €7.6, max: €16.00) and generated an average patient benefit of €12.18 (min: € 2.4, max: € 22.4). In total, €1754.4 were transferred to the *Christoffel-Blindenmission*. Assuming a cost of €30 per eye cataract operation, this amounts to the treatment of about 58 real patients.

3. Results

Data

We consider decisions made by 144 subjects. See Table 1 for the distribution of participants across treatment conditions and degree programs.

Table 1: Overview subjects

Treatment	**Number of subjects**		
	Total	**Medical students**	**Others**
(1) PRIVATE	60	12	48
(2) PUBLIC	60	14	46
(3) REVERSE ORDER (PRIVATE)	24	5	19
Total	144	31	113

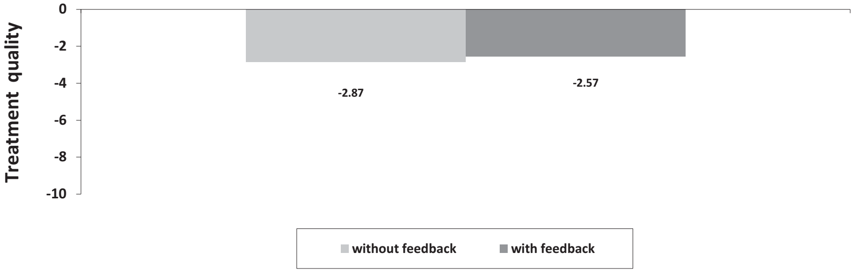
Impact of Feedback Incentives on Treatment Quality

In order to analyze whether feedback incentives serve to improve the quality of medical treatment provided, we first consider the decisions made by all subjects in treatment conditions (1) and (2) and compare their choices in the first task without a feedback incentive to those in the second task with feedback. Treatment quality is defined as the (negative) difference between the optimal benefit a patient can potentially achieve from being treated and the actual benefit he receives from the amount of services he is provided. Average treatment quality thus ranges from 0 (no deviation from optimal quality) to -10.3 (the largest possible average deviation from the optimum across all 9 decisions).

We consider the aggregated decisions made by our subjects for all patients and across all illnesses and degrees of severity, as this best reflects the typical decision situation of a physician who is faced with a heterogeneous group of patients within a time interval such as a month or a quarter.[5]

[5] We control for the impact of the individual illnesses and degrees of severity on the physician's treatment quality in an OLS regression and find significant coefficients for both (see Appendix B.2). This does not detract from our results, as our main concern in this paper is with the general situation of a physician facing a heterogeneous group of patients. However, the impact of feedback incentives on the performance of physicians who deal with more specific sub-populations of patients (e.g. with particularly severe or chronic illnesses) is an interesting subject of further research.

Figure 4: Average treatment quality by task



In the aggregate, the subjects in our experiment provided treatment with a quality of -2.87 on average (so their decisions result in an average loss of patient benefit of 2.87 Taler relative to the optimum; SD = 2.86) in decisions without a feedback incentive, and of -2.57 (SD = 2.74) in decisions with feedback (see Figure 4). This difference is highly statistically significant in a two-sided Mann-Whitney U-test ($p < 0.01$).[6] Our first result is thus:

> *In general, setting a non-monetary feedback incentive for subjects significantly improves the quality of medical treatment they provide to patients.*

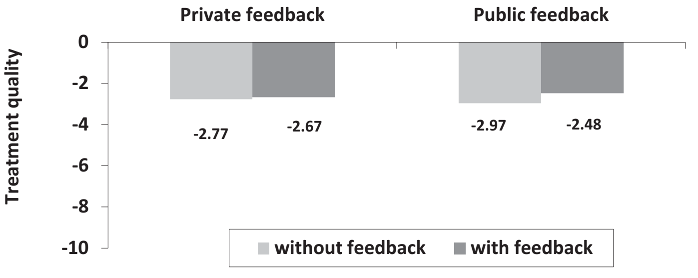
Effect of Feedback Mode

Turning to the relative effects of giving performance feedback privately or publicly, we compare the effect of the feedback incentive across the treatment conditions PUBLIC and PRIVATE. The public feedback incentive in treatment condition (2) led to an improvement in the medical treatment quality from -2.97 (SD = 2.83) to -2.48 (SD = 2.68; see Figure 5).[7] This difference is statistically highly significant ($p < 0.01$). In treatment condition (1), the private feedback incentive improved the average treatment quality slightly from -2.77 (SD = 2.88) to -2.67 (SD = 2.79). This shift is, however, not statistically significant ($p = 0.64$). (The results of these statistical tests are also confirmed in simple OLS regressions; see Appendix B.3.)

[6] Unless noted otherwise, all statistical tests presented here are two-sided Mann-Whitney U-tests and two-tailed Student's t-tests provide very similar results.

[7] Note that while subject behavior in task 1 differs slightly across treatments (1) and (2), this difference is not statistically significant ($p > 0.10$).

Figure 5: Average treatment quality by feedback mode



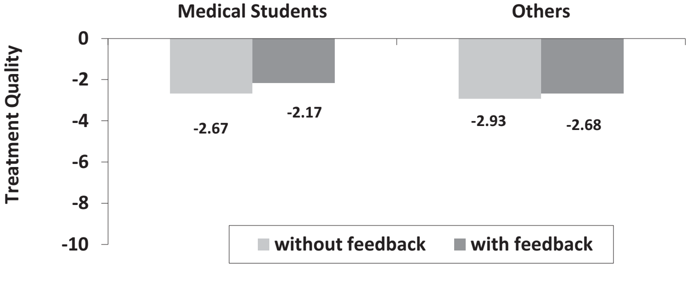
This leads to our second result:

> *The mode in which feedback incentives are provided matters: While public feedback yields a significant improvement in the treatment quality subjects provide, the effect of private feedback is not statistically significant.*

Medical Students

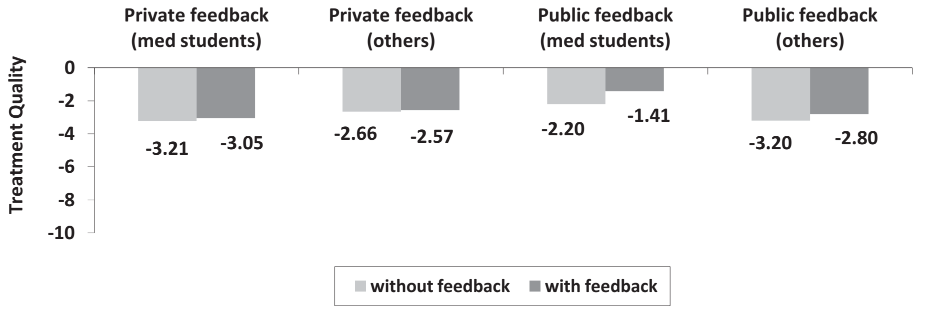
The above two results are generally robust to a relevant subject pool characteristic, whether subjects medical students or not. Considering sub-samples of medical students and other subjects separately, feedback incentives improve average treatment quality from -2.67 (SD = 2.48) to 2.17 (SD = 2.34) for the prior and from -2.93 (SD = 2.95) to -2.68 (SD = 2.83) for the latter (see Figure 6). Both shifts are statistically significant: p = 0.01 and p = 0.07, respectively.

Figure 6: Average treatment quality by degree



The impact of the feedback mode also holds for the two separate sub-samples (see Figure 7): Private feedback tends to improve treatment quality, though the effect is not statistically significant: Medical students in this group achieve a quality of -3.21 (SD = 2.81) without and -3.05 (SD = 2.66) with the incentive, while other subjects improve very slightly from -2.66 (SD = 2.89) to -2.57 (SD = 2.82). For both subsamples, the differences are not statistically significant ($p > 0.70$). The significant effect of the public feedback incentive, on the other hand, is upheld in both groups: Medical students improve their treatment quality from -2.20 (SD = 2.06) to -1.41 (SD = 1.71), while others improve from -3.20 (SD = 2.99) to -2.80 (SD = 2.83); both changes are statistically significant, with $p < 0.01$ and $p < 0.05$, respectively.[8] Moreover, simple OLS regressions show that given public feedback, medical students provide significantly better treatment quality than non-medical students (see Appendix B.4).

Figure 7: Average treatment quality by feedback mode and degree



Hence we find that public feedback significantly improves the treatment quality provided by medical and other students, while the effect of private feedback is not statistically significant for both groups. The effect for public feedback is significantly larger for medical students.

Robustness to Task Order and Subject Characteristics

Using data from the two reverse-order sessions, we find that the results of the private feedback incentive are robust to providing the quality incentive in part 1 of the experiment and not providing it in part 2. Subjects achieve an average treatment quality of -2.38 (SD = 2.52) and -2.13 (SD = 2.45) respectively, which does not represent a statistically significant difference ($p = 0.31$; see Figure 8).

As mentioned above, corresponding controls for a reversed task order are difficult to implement in the public feedback treatment. Making subjects' decisions or their consequences known to other participants in the experiment in part 1 would presumably have an additional influence on the decisions made in part 2, obfuscating the effect of purely reversing the tasks.

[8] While the within-subject effect of public and private feedback is consistent across groups, the treatment quality provided in part one (the non-incentivized task) differs significantly across all pairs of subject groups discussed in this section ($p<0.05$).

We also estimated OLS regressions to control for the influence of subject characteristics (age, gender, family members in the medical profession) and specifics of the decision situation (severity, illness, session, whether subjects knew other participants in the session) on the quality of medical treatment provided in our experiment. None of these factors adds any explanatory power to our analysis (see Appendix B.4).

4. Conclusion

We find feedback as a performance incentive to have an effect on the quality of medical care provided in our experiment. The effect is, however, dependent on the feedback mode: Private feedback has no impact on the quality of care, whereas public feedback has a significant positive impact. Our results are robust to a subject's enrollment in medical education and socio-demographic characteristics as well as to changes in the task order.

So far, there is evidence that physicians react to non-monetary performance incentives (Kolstad, 2013), However, there seems to be little evidence that quality reporting incentives actually lead to better medical treatment quality and lower health care costs (Dranove and Jin, 2010). This may be due to the fact that while treatment quality is typically multidimensional, only some of its aspects can be reported, as e.g. in the US Nursing Home Quality Initiative. In this case, physicians may react to public reporting by improving quality only for the reported measures whilst decreasing quality along non-reported dimensions, for instance by patient selection (Dranove et al., 2003, Werner and Asch, 2005). In our controlled laboratory experiment quality is fully reported. Under these circumstances, we find public feedback incentives to have a positive and significant effect on the quality of medical care provided. Hence, if future policy reforms succeed at establishing more comprehensive ways of reporting quality in health care, this should serve as a tool to increase quality of care.

Our results also suggest that the mode of providing quality feedback is important and should be taken into account by policy makers. The mere motive of boosting self-esteem which underlies private performance feedback does not seem sufficient to align physician interests more closely with patient interests. The additional motive of reputation (image motivation) introduced by public performance feedback, on the other hand, can perhaps foster quality improvement in medical care. Public performance feedback may be a cost-efficient means towards this end – in contrast to monetary pay-for-performance incentives, which also serve to raise patient benefit but are not necessarily cost-efficient (Brosig-Koch et al., 2013b). Future research in this area should be directed towards investigating how monetary mechanisms interact with non-monetary mechanisms, and the conditions under which they enhance or detract from each other.

Literature

Ahlert, M., Felder, S., and B. Vogt (2012): Which Patients Do I Treat?: An Experimental Study with Economists and Physicians, *Health Economics Review*, 2(1), 1–11.

Ariely, D., A. Bracha, S. Meier (2009): Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially, *American Economic Review*, 99, 544–55.

Azmat, G., and N. Iriberri (2010): The importance of relative performance feedback information: Evidence from a natural experiment using high school students, *Journal of Public Economics*, 94(7-8), 435-452.

Baltussen, G., Post, T., van den Assem, M. & Wakker, P.P. (2010): Random Incentive Systems in a Dynamic Choice Experiment, Working Paper, Erasmus University of Rotterdam.

Barankay, I., (2011a): Rankings and social tournaments: evidence from a crowd-sourcing experiment, Working Paper, Wharton School of Business, University of Pennsylvania.

Barankay, I., (2011b): Gender differences in productivity responses to performance rankings: evidence from a randomized workplace experiment, Working Paper. Wharton School of Business, University of Pennsylvania.

Blanes i Vidal, J. and M. Nossol (2011): Tournaments without prizes: evidence from personnel records, *Management Science*, 57 (10), 1721-1736.

Bénabou, R., J. Tirole (2002): Self-confidence and personal motivation, *Quarterly Journal of Economics*, 117(3), 871–915.

Bénabou, R., J. Tirole (2006): Incentives and Prosocial behaviour, *The American Economic Review*, 96, 1652-1678.

Brosig-Koch, J., H. Hennig-Schmidt, N. Kairies, and D. Wiesen (2013a): How to improve patient care? An analysis of capitation, fee-for-service, and mixed incentive schemes for physicians, Ruhr Economic Papers No. 412.

Brosig-Koch, J., H. Hennig-Schmidt, N. Kairies, and D. Wiesen (2013b): How effective are pay-for-performance incentives for physicians? A laboratory experiment, Ruhr Economic Papers No. 413.

Charness, G., Masclet, D. and M.C. Villeval (2011): Competitive preferences and status as an incentive: experimental evidence. CIRANO Working Paper.

Cubitt, R., Starmer, R., Sugden, R. (1998): On the validity of the random lottery incentive system, Experimental Economics, 1(2), 115-131.

Cutler, D.M., Ilckman, R.S. and M. B. Landrum (2004): The Role of Information in Medical Markets: An Analysis of Publicly Reported Outcomes in Cardiac Surgery, *American Economic Review*, 94(2), 342-346.

Dranove, D., Kessler, D., McClellan, M. and M. Satterthwaite (2003): Is More Information Better? The Effects of "Report Cards" on Health Care Providers, *Journal of Political Economy*, 111(3), 555-588.

Dranove, D., and G. Zhe Jin (2010): Quality Disclosure and Certification: Theory and Practice, *Journal of Economic Literature*, 48(4), 935–63.

Ellingsen, T., M. Johannesson (2008): Pride and Prejudice: the Human Side of Incentive Theory, *American Economic Review*, 98, 990-1008.

Eriksson, T., A. Poulsen, M.C. Villeval (2009): Feedback and Incentives: Experimental Evidence, *Labour Economics*, 16, 679-688.

Fischbacher, U. (2007): Z-tree: Zurich toolbox for readymade economic experiments, Experimental Economics, 10(2), 171-178.

Grabowski, D.C., and Town, R.J., (2011): Does information matter? Competition, quality, and the impact of nursing home report cards, *Health Services Research*, 46, 1698–1719.

Greiner, B. (2004): The Online Recruitment System ORSEE 2.0 - A Guide for the Organization of Experiments in Economics, University of Cologne working paper No. 10.

Hannan, E. L., Kilburn, H., Racz, M., Shields, E. and M. R. Chassin (1994): Improving the Outcomes of Coronary Artery Bypass Surgery in New York State, *Journal of the American Medical Association*, 271(10), 761–66.

Hannan, R.L., Krishnan, R., Newman, A. (2008): The effects of disseminating relative performance feedback in tournament and individual performance compensation systems, *Accounts Review*, 83, 893–913.

Hennig-Schmidt, H., R. Selten and D. Wiesen (2011): How Payment Systems Affect Physicians' Provision Behavior? An Experimental Investigation, *Journal of Health Economics*, 30(4), 637-646.

Hibbard, J., Stockard, J. and M. Tusler (2003): Does Publicizing Hospital Performance Stimulate Quality Improvement Efforts?, *Health Affairs*, 22(2), 84–94.

Hibbard, J.H., Stockard, J., and M. Tusler (2005): Hospital performance reports: impact on quality, market share and reputation. *Health Affairs*, 24(4), 1150–1160.

Kolstad, J. T. (2013): Information and quality when motivation is intrinsic: Evidence from surgeon report cards, National Bureau of Economic Research No. w18804.

Kuhnen, C., and A. Tymula (2012): Feedback, self-esteem and performance in organizations, *Management Science,* 58, 94–113.

Laury, S.K. (2006): Pay One or Pay All: Random Selection of One Choice for Payment, Working Paper, Georgia State University.

Mas, A., and E. Moretti (2009): Peers at work, *American Economic Review*, 99(1), 112–145.

Marshall, M.N., Shekelle, P.G., Leatherman, S., and R.H. Brook (2000): The Public Release of Performance Data: What Do We Expect to Gain? A Review of the Evidence, *Journal of the American Medical Association,* 1866–1874.

Maynard, A. (2012): The powers and pitfalls of payment for performance, *Health Economics*, 21 (1), 3-12.

McClellan, M. (2011): Reforming payments to healthcare providers: The key to slowing healthcare cost growth while improving quality?, Journal of Economic Perspectives, 25, 69-92.

Rege, M. and K. Telle (2004): The Impact of Social Approval and Framing on Cooperation in Public Good Situations, *Journal of Public Economics*, 88, 1625-1644.

Rosenthal, G. E., Quinn, L., and D. L. Harper (1997): Declines in hospital mortality associated with a regional initiative to measure hospital performance, *American Journal of Medical Quality*, 12(2), 103-112.

Starmer, C. and R. Sugden (1991): Does the Random-Lottery Incentive System Elicit True Preferences? An Experimental Investigation, American Economic Review, 81(4): 971-978.

Tran, A., and R. Zeckhauser (2012): Rank as an inherent incentive: Evidence from a field experiment, *Journal of Public Economics*, 96(9-10), 645–650.

Werner, R.M., and D. A. Asch (2005): The Unintended Consequences of Publicly Reporting Quality Information, *Journal of the American Medical Association*, 293(10), 1239-1244.

Werner, R.M., Konetzka, R.T., Stuart, E.A., Norton, E.C., Polsky, D., and J. Park (2009): The impact of public reporting on quality of postacute care, *Health Services Research*, 44(4), 1169–87.

Appendix

Appendix A: Instructions and Control Questions

You are participating in an economic experiment on decision behavior. You and the other participants will be asked to make decisions for which you can earn money. Your payoff depends on the decisions you make. At the end of the experiment, your payoff will be converted to Euro and paid to you in cash. During the experiment, all amounts are presented in the experimental currency Taler. 10 Taler equals 8 Euro.

The experiment will take about 90 minutes and consists of two parts. You will receive detailed instructions before each part. Note that none of your decisions in either part have any influence on the other part of the experiment.

Part One

Please read the following instructions carefully. We will approach you in about five minutes to answer any questions you may have. If you have questions at any time during the experiment, please raise your hand and we will come to you.

Part one of the experiment consists of 9 rounds of decision situations.

Decision Situations

In each round you take on the role of a physician and decide on medical treatment for a patient. That is, you determine the quantity of medical services you wish to provide to the patient for a given illness and a given severity of this illness.

Every patient is characterized by one of three illnesses (A, B, C), each of which can occur in three different degrees of severity (x, y, z). In each consecutive decision round you will face one patient who is characterized by one of the 9 possible combinations of illnesses and degrees of severity (in random order). Your decision is to provide each of these 9 patients with a quantity of 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 medical services.

Profit

In each round you are remunerated for treating the patient. Your remuneration increases with the amount of medical treatment you provide. You also incur costs for treating the patient, which likewise depend on the quantity of services you provide. Your profit for each decision is calculated by subtracting these costs from the remuneration.

Every quantity of medical service yields a particular benefit for the patient – contingent on his illness and severity. Hence, in choosing the medical services you provide, you determine not only your own profit but also the patient's benefit.

In each round you will receive detailed information on your screen (see below) on the patient's illness and its severity as well as the remuneration, cost, and patient benefit for each quantity of medical services (see screen shot in Figure 1 above).

Payment

At the end of the experiment one of the 9 rounds of part one will be chosen at random. Your profit in this round will be paid to you in cash.

For this part of the experiment, no patients are physically present in the laboratory. Yet, the patient benefit does accrue to a real patient: The amount resulting from your decision will be transferred to the Christoffel Blindenmission Deutschland e.V., 64625 Bensheim, an organization which funds the treatment of patients with eye cataract.

The transfer of money to the Christoffel Blindenmission Deutschland e.V. will be carried out after the experiment by the experimenter and one participant. The participant completes a money transfer form, filling in the total patient benefit (in Euro) resulting from the decisions made by all participants in the randomly chosen situation. This form prompts the payment of the designated amount to the Christoffel Blindenmission Deutschland e.V. by the University of Duisburg-Essen's finance department. The form is then sealed in a postpaid envelope and posted in the nearest mailbox by the participant and the experimenter.

After the entire experiment is completed, one participant is chosen at random to oversee the money transfer to the Christoffel Blindenmission Deutschland e.V. The participant receives an additional compensation of 5 Euro for this task. The participant certifies that the process has been completed as described here by signing a statement which can be inspected by all participants at the office of the Chair of Quantitative Economic Policy. A receipt of the bank transfer to the Christoffel Blindenmission Deutschland e.V. may also be viewed here.

Comprehension Questions

Prior to the decision rounds we kindly ask you to answer a few comprehension questions. They are intended to help you familiarize yourself with the decision situations. If you have any questions about this, please raise your hand. Part one one if the experiment will begin once all participants have answered the comprehension questions correctly.

Part Two

Please read the following instructions carefully. We will approach you in about five minutes to answer any questions you may have. If you have questions at any time during the experiment, please raise your hand and we will come to you.

Part two of the experiment also consists of 9 rounds of decision situations.

<u>Decision Situations</u>

As in part one of the experiment, you take on the role of a physician in each round and decide on medical treatment for a patient. That is, you determine the quantity of medical services you wish to provide to the patient for a given illness and a given severity of this illness.

Every patient is characterized by one of three illnesses (A, B, C), each of which can occur in three different degrees of severity (x, y, z). In each consecutive decision round you will face one patient who is characterized by one of the 9 possible combinations of illnesses and degrees of severity (in random order). Your decision is to provide each of these 9 patients with a quantity of 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 medical services.

<u>Profit</u>

In each round you are remunerated for treating the patient. Your remuneration increases with the amount of medical treatment you provide. You also incur costs for treating the patient, which likewise depend on the quantity of services you provide. Your profit for each decision is calculated by subtracting these costs from the remuneration.

As in part one, every quantity of medical service yields a particular benefit for the patient – contingent on his illness and severity. Hence, in choosing the medical services you provide, you determine not only your own profit but also the patient’s benefit.

In each round you will receive detailed information on your screen (see below) on the patient’s illness and its severity as well as the remuneration, cost, and patient benefit for each quantity of medical services (see screen shot below).

<u>Payment</u>

At the end of the experiment one of the 9 rounds of part two will be chosen at random. Your profit in this round will be paid to you in cash, in addition to your payment from the round chosen for part one of the experiment.

After the experiment is over, please remain seated until the experimenter asks you to step forward. You will receive your payment at the front of the laboratory before exiting the room.

As in part one, no patients are physically present in the laboratory for part two of the experiment. Yet, the patient benefit does accrue to a real patient: The amount resulting from your decision will be transferred to the Christoffel Blindenmission Deutschland e.V., 64625 Bensheim, an organization which funds the treatment of patients with eye cataract.

The process for the transfer of money to the Christoffel Blindenmission Deutschland e.V. as described for part one of the experiment will be carried out by the experimenter and one participant.

Feedback

In addition to your payment you will receive feedback in this part of the experiment on the quality of treatment you provide as a physician. The best treatment quality is achieved when the patient receives the highest possible benefit. The lower the patient's benefit from the provided amount of services, the worse the treatment quality.

A ranking of all participants in the experiment will be generated. The ranking is based on the treatment quality provided in the decision situation chosen for payment in this part of the experiment. The participant with the highest treatment quality ranks first, the participant with the worst treatment quality ranks last. Participants with equal treatment quality share ranks.

[Private feedback treatment:]

You will see your placement in this ranking on your screen at the end of the experiment. Every participant only learns their own rank, not those of other participants.

[Public feedback treatment:]

This ranking will be shown on your screen once the experiment has been completed. A member of the laboratory staff will then ask all participants to stand up. The ranking will be read out aloud. (The participants' ranks and seat number will be stated, not their names or specific decisions.) When your seat number is called, please hold up the sign with the number so that it is visible to all participants.

We kindly ask you to not talk to anyone about the content of this session in order to prevent influencing other participants after you. Thank you for your Collaboration!

Exemplary Comprehension Question Part 1:

Quantity of medical treatment	Fee-for-service (in Taler)	Costs (in Taler)	Profit (in Taler)	Benefit of the patient with illness F and severity y (in Taler)
0	0.00	0.00	0.00	15.00
1	4.00	0.20	3.80	16.00
2	8.00	0.80	7.20	17.00
3	12.00	1.80	10.20	18.00
4	16.00	3.20	12.80	19.00
5	20.00	5.00	15.00	20.00
6	24.00	7.20	16.80	19.00
7	28.00	9.80	18.20	18.00
8	32.00	12.80	19.20	17.00
9	36.00	16.20	19.80	16.00
10	40.00	20.00	20.00	15.00

Assume that a physician wants to provide 2 quantities of medical treatment for the patient depicted above.

a) What is the fee-for-service?
b) What are the costs?
c) What is the profit?
d) What is the patient benefit?

Appendix B: Further Tables

B.1 Decision Parameters

		Quantity (q)										
Treatment	Variable	0	1	2	3	4	5	6	7	8	9	10
all	$R_{kl}^{Part\ 1}$	0	2	4	6	8	10	12	14	16	18	20
	$R_{kl}^{Part\ 2}$	0	2	4	6	8	10	12	14	16	18	20
all	c_{kl}	0	0.1	0.4	0.9	1.6	2.5	3.6	4.9	6.4	8.1	10
all	π_{kl}^{Part1}	10	9.9	9.6	9.1	8.4	7.5	6.4	5.1	3.6	1.9	0
	π_{kl}^{Part2}	10	9.9	9.6	9.1	8.4	7.5	6.4	5.1	3.6	1.9	0
all	B_{Ax}	4	5	6	7	6	5	4	3	2	1	0
	B_{Ay}	2	3	4	5	6	7	6	5	4	3	2
	B_{Az}	0	1	2	3	4	5	6	7	6	5	4
	B_{Bx}	7	8	9	10	9	8	7	6	5	4	3
	B_{By}	5	6	7	8	9	10	9	8	7	6	5
	B_{Bz}	3	4	5	6	7	8	9	10	9	8	7
	B_{Cx}	8	10	12	14	12	10	8	6	4	2	0
	B_{Cy}	4	6	8	10	12	14	12	10	8	6	4
	B_{Cz}	0	2	4	6	8	10	12	14	12	10	8

B.2 Control for Illnesses and Severities (OLS regression, aggregated data)

VARIABLES	Aggregated
Severity	0.647***
	(0.039)
Illness	-1.013***
	(-0.0737)
Constant	-3.928***
	(0.241)
Observations	2,160
R-squared	0.229
N_clust	120

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

B.3 Effect of Feedback Mode (OLS regression)

VARIABLES	Private	Public
Feedback Incentive	0.106	0.491***
	(0.0808)	(-0.0989)
Constant	-2.774***	-2.967***
	(0.23)	(0.253)
Observations	1,080	1,080
R-squared	0.000	0.008
N_clust	60	60

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

B.4 Control for Subject Characteristics

B.4.1 Descriptive Subject Characteristics (excluding reverse-order subjects):

Variable	Average (120 subjects)	Min	Max
Female	0.6	0	1
Age	23.5	18	49
Medical student	0.142	0	1
Econ student	0.325	0	1
Parents in health care job	0.2	0	1
Knew other subjects in session	0.31	0	3

B.4.2 Regressions (OLS, clustered by subjects, excluding reverse-order subjects):

Aggregate Data

VARIABLES	1	2	3	4	5
Feedback Incentive	0.298***	0.298***	0.298***	0.298***	0.298***
	(0.066)	(0.066)	(0.066)	(0.066)	(0.066)
Age	0.00729				
	(0.0467)				
Female		0.41			
		(0.357)			
Parents in health care job			-0.291		
			(0.463)		
Medical student				0.389	
				(0.371)	
Knew other subjects in session					-0.434
					(0.342)
Constant	-3.042***	-3.117***	-2.812***	-2.955***	-2.736***
	(1.091)	(0.285)	(0.186)	(0.201)	(0.18)
Observations	2,160	2,160	2,160	2,160	2,160
R-squared	0.003	0.008	0.005	0.006	0.011
N_clust	120	120	120	120	120

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Private Feedback

VARIABLES	1	2	3	4	5
Feedback Incentive	0.106	0.106	0.106	0.106	0.106
	(0.0808)	(0.0808)	(0.0808)	-0.0808	(0.0808)
Age	0.134*				
	(0.0703)				
Female		0.0243			
		(0.474)			
Parents in health care job			(0.849)		
			(0.537)		
Medical student				0.51	
				(0.586)	
Knew other subjects in session					-0.940*
					(0.514)
Constant	-5.819***	-2.788***	-2.533***	-2.672***	-2.524***
	(1.635)	(0.346)	(0.258)	(0.257)	(0.227)
Observations	1,080	1,080	1,080	1,080	1,080
R-squared	0.023	0	0.019	0.006	0.025
N_clust	60	60	60	60	60

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Public Feedback

VARIABLES	1	2	3	4	5
Feedback Incentive	0.491***	0.491***	0.491***	0.491***	0.491***
	(0.099)	(0.099)	(0.099)	(0.099)	(0.099)
Age	-0.0322				
	(0.0488)				
Female		0.823			
		(0.548)			
Parents in health care job			0.772		
			(0.859)		
Medical student				1.194***	
				(0.415)	
Knew other subjects in session					-0.168
					(0.465)
Constant	-2.185*	-3.488***	-3.057***	-3.245***	-2.908***
	(1.174)	(0.465)	(0.264)	(0.308)	(0.278)
Observations	1,080	1,080	1,080	1,080	1,080
R-squared	0.012	0.028	0.016	0.041	0.009
N_clust	60	60	60	60	60

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1