

Stress in performance-related pay: the effect of payment contracts and social-evaluative threat

Nicole Andelic, Julia Allan, Keith A. Bender, Daniel Powell & Ioannis Theodossiou

To cite this article: Nicole Andelic, Julia Allan, Keith A. Bender, Daniel Powell & Ioannis Theodossiou (2023) Stress in performance-related pay: the effect of payment contracts and social-evaluative threat, *Stress*, 26:1, 2283435, DOI: [10.1080/10253890.2023.2283435](https://doi.org/10.1080/10253890.2023.2283435)

To link to this article: <https://doi.org/10.1080/10253890.2023.2283435>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 27 Nov 2023.



Submit your article to this journal [↗](#)



Article views: 86



View related articles [↗](#)



View Crossmark data [↗](#)

Stress in performance-related pay: the effect of payment contracts and social-evaluative threat

Nicole Andelic^a, Julia Allan^b, Keith A. Bender^a, Daniel Powell^c and Ioannis Theodossiou^a

^aEconomics Department and Centre for Labour Market Research, Business School, University of Aberdeen, Aberdeen, United Kingdom;

^bPsychology, University of Stirling, Stirling, United Kingdom; ^cHealth Psychology, Institute of Applied Health Sciences, University of Aberdeen, Aberdeen, United Kingdom

ABSTRACT

There is some evidence that performance-related pay (PRP) leads to higher levels of stress as it incentivizes employees to work harder for longer. However, PRP in the workplace also typically involves performance monitoring, which may introduce an additional source of stress via social-evaluative threat (SET). The current study examined the effect of PRP on stress while varying the level of performance monitoring/SET. Using an incentivized mixed design experiment, 206 participants completed a simulated work task after being randomly allocated to either a PRP contract (£0.20 per correct response, $n=110$) or minimum-performance fixed payment contract (£5 for ≥ 10 correct responses; £0 for < 10 , $n=96$) condition. All participants completed the task during a high SET (explicit performance monitoring) and low SET (no monitoring) condition. Subjective and objective stress were measured through self-report and salivary cortisol. High SET led to higher levels of self-reported stress but not cortisol, whereas there was no effect of the payment condition on either self-reported stress or cortisol. A statistically significant interaction revealed that high SET-fixed payment participants were significantly more stressed than those in the high SET-PRP group. Estimating the regressions separately for high- and low-performing individuals found that the effect was driven by low-performing individuals. These results suggest that fixed payment contracts that have a minimum performance threshold and which include performance monitoring and SET can be more stressful than traditional piece-rate PRP contracts. The current study suggests that incorporating performance monitoring and SET into payment contracts may affect the well-being of employees.

ARTICLE HISTORY

Received 18 February 2023

Accepted 28 October 2023

KEYWORDS

Stress; cortisol; payment contracts; social-evaluative threat; performance-related pay; acute stress

1. Introduction

Performance-related pay (PRP) is a payment structure where employees' pay is based fully or partially on their performance, as opposed to a payment that is fixed regardless of work performance. As employees paid by PRP are highly motivated to produce more to increase their earnings, PRP is typically linked with advantages such as higher productivity (Lazear, 2000). However, there is a growing body of literature also linking PRP with poor health outcomes, including poorer mental health (Dahl and Pierce, 2020), more cardiovascular and digestive issues (Bender & Theodossiou, 2014), higher systolic blood pressure and higher levels of fibrinogen (Andelic et al., 2023). Although there may be multiple pathways through which PRP can lead to poor health, one potential mechanism is through working at a higher rate or working long hours which leads to higher levels of physiological stress.

Physiological changes in response to stressors at work are important as they can have long-term effects. Indeed, persistent stress is associated with higher inflammation in the

body (Aschbacher et al., 2013; Rohleder, 2014) and can lead to the progression of illnesses (McEwen, 1998). Analysis of a large-scale survey found higher systolic blood pressure and higher levels of fibrinogen among PRP employees (Andelic et al., 2023), both of which are associated with chronic stress, and there is some experimental evidence suggesting that PRP can lead to higher levels of stress; Allan et al. (2021) measured salivary cortisol, a marker of stress, in participants after they had completed a simulated work (calculation) task under both a PRP and a fixed payment condition. In the PRP condition, participants received £0.20 per correct calculation, whereas they earned a fixed payment of £5 if they completed a minimum of 10 correct calculations in the fixed payment condition. The study found that participants had higher levels of cortisol directly after the PRP condition than after the fixed payment condition, suggesting that employees who are incentivized through PRP may experience higher levels of acute physiological stress.

However, Allan et al. (2021) automatically tracked the correct answers on which payment was based. In contrast, the

CONTACT Nicole Andelic  nicole.andelic@abdn.ac.uk  Economics Department, Business School, University of Aberdeen, Aberdeen, United Kingdom

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

performance of PRP employees in the workplace is routinely monitored by supervisors and peers. In any job involving monitoring, there is potential for a social-evaluative threat (SET). SET occurs when there is a perceived threat to your identity through social evaluation, such as being negatively judged by others due to a trait or inability to perform a task, which in turn increases cortisol reactivity. This response is particularly heightened when there is an element of motivated performance (see meta-analysis on SET and acute physiological stress by Dickerson & Kemeny, 2004), i.e. when there is a task with a clear goal or incentive. Importantly, for PRP employees, performance in the workplace may lead to stress both through SET and via the direct impact it has on work hours and income.

The current study used a similar design to Allan et al. (2021) to examine the effect of PRP on acute physiological stress, but with the addition of a high SET condition where explicit performance monitoring was incorporated. In line with Allan et al. (2021) and the wider literature on PRP and stress, we hypothesized that PRP participants would show higher levels of subjective stress and cortisol. In line with Dickerson and Kemeny (2004), we also hypothesized that subjective stress and cortisol responses would be higher in the high SET condition than the low SET condition. Finally, we hypothesized that there would be an interaction effect and that the cortisol response would be higher in a condition with both high SET and high levels of motivated performance, such as PRP. Consequently, we expected both subjective and physiological stress to be higher in a combined high SET-PRP condition in comparison to low SET-PRP, high SET-fixed payment, or low SET-fixed payment.

2. Materials and methods

2.1. Design

The study consisted of two financially incentivized sessions where participants completed a simulated work task (mathematical calculations) in a computer lab. Sessions were conducted one week apart to measure the impact of payment contracts and SET on both self-reported and physiological stress (cortisol). The experiment was a 2×2 mixed factorial design with a between-subjects factor (payment contract) and a within-subjects factor (SET). Payment contracts had two levels: PRP or fixed. All participants were randomly assigned to conditions. The PRP contract paid £0.20 per correct answer; the fixed payment contract was a minimum performance contract with £5 paid where 10 or more correct answers were given and £0 where this was not achieved. The SET condition had two levels: high SET or low SET. The order was counterbalanced. High SET involved the active monitoring of the participant's performance on a public leaderboard showing their relative performance and with a supervisor monitoring performance in the room. A leaderboard was chosen as a league table by performance, which is often used in workplace settings such as sales or call center settings (Bain & Taylor, 2000). The low SET condition had no leaderboard and no visible monitoring of performance in

the room, but payment was still tracked, as in Allan et al. (2021). Ethical approval was granted by the University of Aberdeen, School of Medicine, Medical Sciences and Nutrition (CERB/2021/3/2058).

2.2. Participants

The participants consisted of students at the University of Aberdeen. Sample size calculations were carried out in Gpower 3.1 (Faul et al., 2007). Previous research (Allan et al., 2021) found small-to-moderate effects of PRP on cortisol ($d=0.28$) and moderate effects of PRP on self-reported stress ($d=0.44$). Research comparing low and high SET found moderate-to-large effects ($\eta^2 = 0.10$, Bosch et al., 2009). The current study estimated a minimum sample size of 88 participants per independent group (a total of 176) was needed using a moderate effect size, 80% power and α of 0.05%.

Two-hundred and fifteen participants took part. Out of these, nine participants had salivary cortisol levels that were more than 4 SDs from the mean. These were excluded as recommended by Nicolson (2008), leaving 206 participants. An additional six participants did not complete all survey items. To maximize the sample size, they were retained for the t -test analyses where possible, but not for the multiple linear regressions, which required survey responses to be included as control variables.

2.3. Materials

2.3.1. Work task

The simulated "work task" was an incentivized math calculation task on the computer. Participants were asked to complete as many math calculations as possible in 10 min. Time left to complete the task was visible at the top right corner of the screen. Participants could complete up to a maximum of 50 possible calculations consisting of 10 addition, 30 multiplication, and 10 division questions presented in the same order for all participants. The exact values used for these calculations were randomly generated for each individual but within restricted parameters (addition and multiplication solutions had to be within 0–10 000, division solutions had to be within 0–100) to maintain a similar level of difficulty despite the randomization. Participants could not skip questions but they were not penalized for incorrect answers. They were given instant feedback after each calculation and the total number of correct responses was displayed at the top of the screen during the full task in such a way that only the participant could see or, in the case of the SET condition, that a monitoring supervisor stood immediately behind them could see.

2.3.2. Questionnaires

Participants completed two surveys during each session. The first survey was carried out before the work task and consisted of the GHQ-12 (general health questionnaire-12; Goldberg & Williams, 1988) as a measure of general stress and a single-item measure asking them to rate their arithmetic skill from 1 ("very good") to 5 ("very poor").

The second survey took place after the completion of the work task and included six single-item measures designed to assess different aspects of task-related stress and perceptions of evaluation. The first four items were used by Allan et al. (2021): “After the task, how stressed do you feel?” “After the task, how exhausted do you feel?” “How much effort did you exert solving the mathematical problems in the previous 10 min?” and “Did you feel under strain when solving the mathematical problems in the previous 10 min?” An additional two items were included to gain insight into self-reported SET: “After the task, do you feel at risk of being negatively evaluated by others?” and “Do you feel that others will believe that your performance reflects your ability?” All six items were rated from 1 (“None at all”/“Not at all”) to 4 (“Much more”/“Great”). A fifth option allowed participants to choose “Care not to answer” to mimic the structure of the GHQ-12.

Participants also rated their arithmetic ability again and self-reported their sex at birth, age, year, and subject of study, as well as whether they had taken part in a similar experiment previously. They were then presented with a checklist of activities and medications that could affect cortisol levels (e.g. smoking, eating within the last hour, etc.) and asked to indicate all items that applied to them.

2.3.3. Cortisol

Cortisol was measured from salivary samples collected using Cortisol Salivettes® (Sarstedt, Nümbrecht, Germany) four times throughout the experiment (details below). To control for the diurnal patterns of cortisol, all experiments took place at 2 p.m. Participants were asked to avoid eating, drinking coffee, smoking nicotine, sleeping, brushing their teeth, or doing vigorous exercise in the two hours immediately before the experiment. All samples were frozen at -23°C until all data collection was complete. All samples were analyzed by Trier Biochemischen Labor in duplicate using the DELFIA assay (a time-resolved fluorescent immunoassay). The intra-assay coefficient of variation was 5.08%.

2.4. Procedure

Data collection took place from October 2021 to March 2022 in the Scottish Experimental Economics Laboratory at the University of Aberdeen. At the time, COVID-19 restrictions were in place for in-person experiments. Consequently, we implemented a staggered arrival for participants, a one-way system and social distancing in the lab during all sessions. All participants completed a health and temperature check upon arrival and were asked to wear a mask whenever entering/exiting the lab. Staff running the experiments wore full personal protective equipment (PPE). Up to twelve participants could take part in the lab at the same time with social distancing.

Upon arrival at 2 p.m., participants were randomly assigned to a seat in the lab. After they were seated, they were given further information about the session. To allow for the 10-min it takes for cortisol levels to be reflected in the saliva (Nicolson, 2008), participants relaxed for 10-minutes during which time they could use coloring-in sheets if they wanted to (see Figure 1) before providing the baseline cortisol sample. This was followed by the completion of the first survey. After this, they were provided with three practice math calculations and unlimited time to complete them.

Participants were then randomly allocated into a payment contract and SET-condition and asked to complete the simulated work task.

After the work task, participants were told the total amount earned from the work task, and they then provided another saliva sample and completed the second survey. After the second survey, participants completed two more 10-min relaxation phases, each followed by a third and fourth cortisol sample. The following week, they returned to do the same protocol again, but this time they were allocated to the other SET condition.

2.5. Statistical analysis

Baseline cortisol levels vary from person to person and may even fluctuate between days (Nicolson, 2008). Consequently,

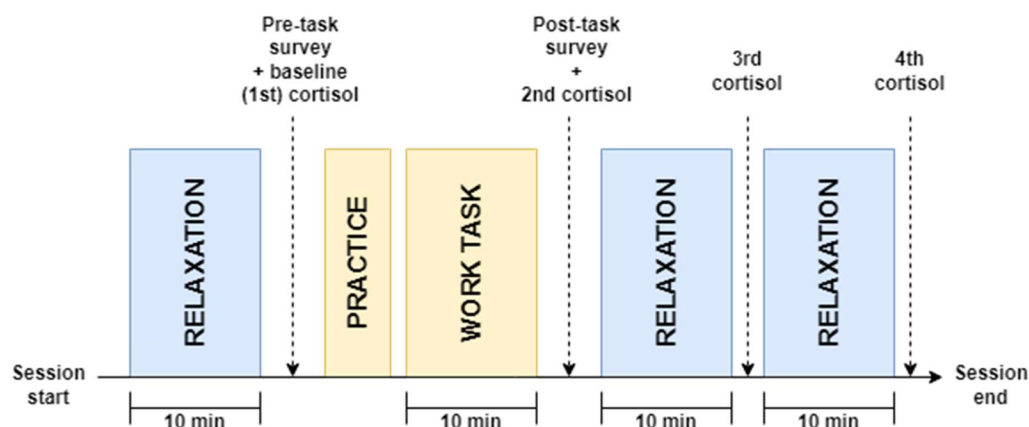


Figure 1. Experimental procedure. A timeline presenting a 10-min block of relaxation, followed by a block of practice (with no time specified), a 10-min block of work task and two more 10-min blocks of relaxation. Arrows indicate cortisol collection points after each timed block and survey completion after the first relaxation phase and after the work task.

we were interested in the change between pre-task and post-task cortisol rather than overall cortisol output during that period. Three cortisol variables measuring the change in cortisol were calculated: (1) area under the curve with respect to increase (AUCi); (2) the difference between baseline and peak levels of cortisol post-task (peak change); and (3) the difference between the baseline and final cortisol sample (overall change). For all computed cortisol variables, a higher score indicates higher stress reactivity. For completeness, we also report the overall measure of cortisol output, area under the curve with respect to ground (AUCg), although we did not expect any group differences. The GHQ score was computed using the Likert scoring method (0–3).

Independent and paired samples t-tests were used to compare scores on each self-reported stress item and cortisol level between payment contracts and SET conditions, respectively. To test the hypothesis that PRP-SET would lead to higher physiological and self-reported stress, group differences in both physiological and self-reported stress between SET and payment conditions were tested using mixed-effects linear regressions (random intercept) while controlling for sample characteristics (sex at birth, age group, higher education level, prior participation in other experiments, and order of SET) and potential cortisol confounders (any medications or activities disclosed in the confounder checklist were coded as 0=no confounders; 1=one or more confounders). Finally, an interaction effect between SET and PRP was examined by estimating a mixed effects linear regression (random intercept) with all the previously listed controls and predicting stress (either self-reported or cortisol) from the experimental condition. All analyses were carried out in R using the base package. For all tests, α was set at 0.05, and effect sizes were examined with 95% confidence intervals.

3. Results

3.1. Descriptive statistics

Out of the 206 participants, 110 were randomly allocated to the PRP group, and 96 were allocated to the fixed payment group. Sample characteristics are presented in Table 1. Approximately half—53% of the PRP contract participants and 52% of the fixed contract participants—took part in the

Table 1. Sample characteristics in first session.

| | PRP | Fixed payment |
|---------------------------------|-----------------|-----------------|
| | <i>n</i> = 110 | <i>n</i> = 96 |
| Female | 70 (63.64%) | 63 (65.63%) |
| 24+ years | 35 (31.82%) | 24 (25.00%) |
| Undergrad | 89 (80.91%) | 81 (84.38%) |
| Edu—arts | 33 (30.00%) | 32 (33.33%) |
| Edu—business school | 31 (28.18%) | 26 (27.08%) |
| Edu—life sciences | 24 (21.81%) | 23 (23.96%) |
| Edu—physical sciences | 21 (19.09%) | 14 (14.58%) |
| GHQ-12 score | <i>M</i> = 1.12 | <i>M</i> = 1.15 |
| Arithmetic skill pre-task score | <i>M</i> = 2.47 | <i>M</i> = 2.57 |

GHQ-12 = General Health Questionnaire (Goldberg & Williams, 1988). GHQ-12 scores range from 0–3 with scores above 2 generally considered to be indicative of poor mental health (Goldberg et al., 1997). Arithmetic skill scored from 1 (“very good”) to 5 (“very poor”).

high SET condition first. Fifteen participants did not return for their second session, and six participants did not complete all the self-reported item ratings. To maximize the sample size, they have been included in analyses where possible.

3.2. Self-reported outcome measures

Compared with those in the low SET condition, participants in the high SET condition reported significantly higher levels of self-reported stress ($t(189) = 4.58, p < .001, d = 0.40$), feeling more at risk of being negatively evaluated by others ($t(188) = 3.50, p = .001, d = 0.26$), more exhaustion ($t(187) = 3.02, p = .003, d = 0.22$), greater effort ($t(189) = 2.51, p = .013, d = 0.18$) and higher strain ($t(189) = 3.13, p = .002, d = 0.23$) (see Table 2). There was no significant difference between groups in beliefs that others would believe that their performance reflected their ability ($t(188) = 1.68, p = .094, d = 0.11$) or self-rated arithmetic skill after the task ($t(189) = 1.54, p = .126, d = 0.13$).

There were few differences between the PRP and fixed payment groups. Compared with the fixed payment contract group, the PRP group were not significantly more stressed ($t(394) = 0.50, p = .617, d = 0.05$), exhausted ($t(392) = 1.00, p = .320, d = 0.10$), or strained ($t(394) = 1.11, p = .269, d = 0.11$), and did not feel more at risk of being negatively evaluated ($t(393) = 0.28, p = .783, d = 0.03$), feel that others would believe their performance reflected their ability ($t(393) = 1.01, p = .314, d = .10$) or perceive their arithmetic skill as poorer ($t(394) = -0.54, p = .590, d = -0.05$), although they did rate themselves as putting in more effort ($t(394) = 2.58, p = .010, d = 0.26$).

In the mixed effects model examining PRP and SET together, there was again no significant difference in self-reported stress between the PRP and fixed payment groups. However, participating in the high SET condition was a significant predictor of higher stress ($\beta = 0.41, p < .001$) even

Table 2. Self-reported means and standard deviations for each group.

| | PRP | | Fixed payment | |
|---|----------------|----------------|----------------|----------------|
| | Low SET | High SET | Low SET | High SET |
| Stress | 1.26 (0.79) | 1.52 (0.84) | 1.13 (0.87) | 1.55 (0.86) |
| Exhaustion | 1.02 (0.73) | 1.07 (0.69) | 0.82 (0.76) | 1.12 (0.81) |
| Effort | 2.19 (0.72) | 2.35 (0.61) | 2.04 (0.73) | 2.14 (0.71) |
| Strain | 1.71 (0.89) | 1.87 (0.82) | 1.56 (1.03) | 1.82 (0.82) |
| Risk of being negatively judged by others | 0.87 (1.09) | 1.12 (1.15) | 0.86 (1.08) | 1.18 (1.12) |
| Risk that others would believe that performance reflected ability | 1.52 (1.11) | 1.61 (1.11) | 1.60 (1.03) | 1.74 (1.04) |
| Arithmetic ability | 2.56 (0.99) | 2.72 (1.00) | 2.64 (1.10) | 2.74 (1.00) |

SET = social evaluative threat; PRP = performance-related pay. Stress, exhaustion, effort and strain refer to the first four self-reported items in the post-task survey: “After the task, how stressed do you feel?” “After the task, how exhausted do you feel?” “How much effort did you exert solving the mathematical problems in the previous 10 minutes?” and “Did you feel under strain when solving the mathematical problems in the previous 10 minutes?” (all scored from 0 = “None at all”/“Not at all” to 3 = “Much more”/“Great”).

after controlling for age, sex, level of study, prior participation and order of session. There was no significant interaction between PRP and SET ($\beta = -0.18, p = .212$).

3.3. Cortisol outcome measures

Similar to the findings of Allan et al. (2021), the means of all cortisol change variables were negative, suggesting that anticipation of taking part in the task was more stressful than the work task itself (see Table 3). Consequently, any difference discussed here is more appropriately labeled as cortisol recovery rather than stress reactivity. Regardless, there was no significant difference between the high and low SET condition for AUCi ($t(190) = 0.46, p = .323, d=0.05$), peak change ($t(190) = 1.13, p = .129, d=0.11$) or overall change ($t(190) = 0.22, p = .414, d=0.02$). Similarly, there was no significant difference for AUCg ($t(395) = -1.21, p = .227, d=0.12$), peak change ($t(395) = -1.41, p = .160, d=0.14$) or overall change ($t(395) = -1.54, p = .123, d=0.16$) when comparing the PRP and fixed payment contracts.

There was no significant difference between the PRP and fixed payment groups in cortisol AUCg ($t(395) = -0.20, p = .840, d=0.02$), but AUCg did differ significantly between the high and low SET condition ($t(190) = 2.41, p = .017, d=0.16$), suggesting that although it did not manifest as a larger cortisol change, the overall cortisol output was higher during the high SET condition.

A mixed effects model was carried out to examine if PRP or SET (or their interaction) predicted cortisol change after including control variables. As can be seen in Table 4, there was a statistically significant interaction between PRP and SET for all three cortisol change outcomes, AUCi ($\beta = -15.20, p = .047$), peak change ($\beta = -0.58, p = .042$), and overall change ($\beta = -0.81, p = .017$) such that, where SET was high, recovery was slower in the fixed payment group than the PRP (piece-rate) group (see Figure 2).

Table 3. Cortisol means and standard deviations for each group.

| | PRP | | Fixed payment | |
|------------------------------|-------------------|-------------------|-------------------|-------------------|
| | Low SET | High SET | Low SET | High SET |
| AUCi | -23.25 (46.07) | -28.81 (54.46) | -25.54 (41.98) | -14.67 (46.29) |
| Peak change | -0.33 (1.63) | -0.43 (1.76) | -0.39 (1.48) | 0.14 (2.13) |
| Overall change | -1.27 (1.93) | -1.65 (2.36) | -1.37 (1.77) | -0.88 (2.25) |
| AUCg | 148.85 (85.84) | 159.26 (85.78) | 148.46 (70.21) | 162.52 (68.65) |
| Cortisol 1st sample (nmol/L) | 4.30 (2.73) | 4.70 (2.97) | 4.35 (2.36) | 4.43 (2.23) |
| Cortisol 2nd sample (nmol/L) | 3.73 (2.35) | 4.06 (2.27) | 3.71 (1.84) | 4.01 (1.77) |
| Cortisol 3rd sample (nmol/L) | 3.48 (2.02) | 3.60 (2.13) | 3.44 (1.76) | 4.03 (1.97) |
| Cortisol 4th sample (nmol/L) | 3.03 (1.66) | 3.05 (1.52) | 2.98 (1.47) | 3.55 (1.88) |

SET = social evaluative threat; PRP = performance-related pay; AUCi = area under the curve with respect to increase; AUCg = area under the curve with respect to ground.

As the difficulty, and therefore stress, of the work task may have varied between those with weak and strong mathematical skills, further sub-group analyses were carried out in which the participants were split into relatively low- and relatively high-performers using a median split from their number of correct calculations during the simulated work task. Carrying out the same mixed effects model separately for the low and high performers revealed that the interaction pattern predicting overall change was only evident among the low performing participants subgroup ($\beta = -1.22, p = .015$) and not among the high performing subgroup ($\beta = -0.29, p = .516$). Similarly, there was no significant interaction when predicting AUCi ($\beta = -11.93, p = .316$) or peak change ($\beta = -0.41, p = .340$) among high performers. Although there were also no statistically significant interactions when predicting AUCi ($\beta = -18.23, p = .070$) or peak change ($\beta = -0.75, p = .055$) among low performers, the direction of these effects are consistent with the results for the total sample.

4. Discussion

In contrast to our hypotheses, we only found a significant effect of SET and no effect of PRP on cortisol responses. Furthermore, an interaction effect revealed that high SET led to higher levels of cortisol when paired with the minimum performance fixed payment contract compared with when paired with PRP. Among the control variables, we found that older participants and participants who had participated in previous experiments in the lab were more likely to be stressed or recover at a slower pace. The relationship between slow recovery and prior experience is likely to be at least in part due to these participants feeling less anticipation prior to the work task and therefore having lower baseline levels of stress. Importantly, the randomized allocation used resulted in a comparable proportion (roughly two-thirds) of participants in both the PRP and fixed payment conditions having had prior experience taking part in a lab experiment.

The small but significant effect of SET on cortisol is well-established in the literature (Dickerson & Kemeny, 2004). The relatively small size of the effect here is not surprising; the nature of the experiment meant that performance had to be monitored to some extent in order to calculate the payment, regardless of the condition. Consequently, there was some degree of SET in both of the experimental conditions. By including a leaderboard, there was a threat of evaluation from peers as well as the supervisor, and by instructing the experimenter to be present in the room during the work task, SET was made more salient in the high SET condition. The difference in self-reported threat seen in the current study suggests that this manipulation worked as intended.

The absence of a statistical difference in stress between the PRP and fixed payment conditions is more surprising considering the findings of Allan et al. (2021). However, Dickerson and Kemeny (2004) find that motivated (e.g. incentivized) performance increases cortisol reactivity when paired with uncontrollability and SET, but not on its own. Interestingly, examining the difference in self-reported stress between PRP and fixed payment in just the low SET conditions shows

Table 4. Mixed effects regressions for cortisol change outcomes by experimental group.

| | AUCi | | | Peak change | | | Overall change | | |
|-----------------------------------|----------------|---------------|-----------------|--------------|-------------|-----------------|----------------|-------------|-----------------|
| | <i>B</i> | <i>SE</i> | <i>p</i> | β | <i>SE</i> | <i>p</i> | β | <i>SE</i> | <i>p</i> |
| Fixed effects | | | | | | | | | |
| (Intercept) | -40.59 | 10.57 | <.001 | -0.57 | 0.39 | .146 | -2.20 | 0.46 | <.001 |
| PRP | -0.43 | 6.35 | .946 | -0.04 | 0.24 | .862 | 0.01 | 0.28 | .978 |
| High SET | 10.47 | 5.48 | .058 | 0.51 | 0.21 | .014 | 0.49 | 0.24 | .047 |
| PRP*high SET | -15.20 | 7.59 | .047 | -0.58 | 0.29 | .042 | -0.81 | 0.34 | .017 |
| 24+ years | 20.46 | 6.83 | .003 | 0.59 | 0.25 | .021 | 0.93 | 0.30 | .002 |
| Undergraduate | -6.36 | 8.16 | .437 | -0.57 | 0.30 | .058 | -0.05 | 0.36 | .887 |
| Male | 10.12 | 5.55 | .070 | 0.37 | 0.20 | .068 | 0.26 | 0.24 | .280 |
| Second session | 3.16 | 3.81 | .408 | 0.14 | 0.14 | .320 | 0.17 | 0.17 | .321 |
| Confounding activities | -8.48 | 5.56 | .128 | -0.28 | 0.21 | .182 | -0.50 | 0.25 | .042 |
| Confounding medications | 4.43 | 6.50 | .496 | -0.002 | 0.24 | .994 | 0.17 | 0.29 | .551 |
| Prior participation | 16.62 | 5.65 | .001 | 0.57 | 0.21 | .007 | 0.77 | 0.25 | .002 |
| Random effects (variances) | | | | | | | | | |
| Residual | 1407.64 | 144.90 | <.001 | 2.00 | 0.21 | <.001 | 2.78 | 0.28 | <.001 |
| Subject (Intercept) | 617.28 | 155.05 | <.001 | 0.78 | 0.21 | <.001 | 1.15 | 0.29 | <.001 |

SET = social evaluative threat; PRP = performance-related pay. Reference categories are fixed payment performance, low SET, 18–23 years, postgraduate, female, first session, no confounding activities, or medications and no prior participation. Bolded variables indicate statistical significance ($\alpha = .05$).

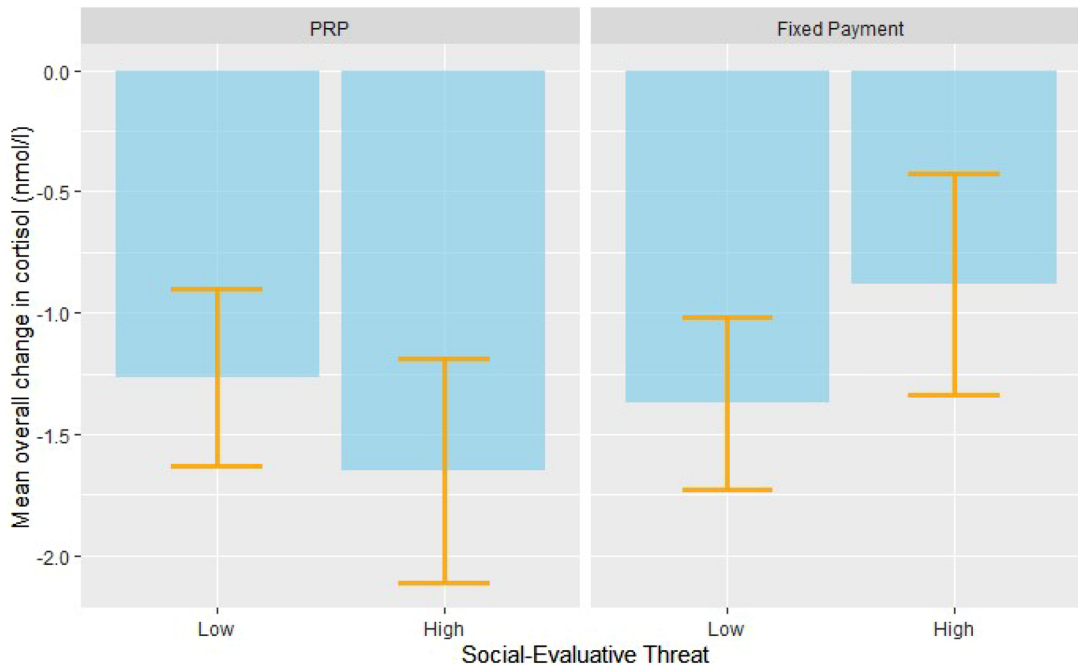


Figure 2. Mean overall cortisol change levels per subgroup with 95% CI. A boxplot with four bars demonstrating the mean overall change for each payment group-SET combination. The fixed payment-high SET bar has the highest level of overall mean change whereas the PRP-high SET bar has the lowest overall change.

similar trends as found in previous studies (Allan et al., 2021). The current study employs a smaller sample, and the current design measures SET within-participants to allow for a repeated measures design across SET instead of across payment contracts, which may explain why the difference is not statistically significant. However, under conditions of high SET, this trend is reversed, and fixed payments lead to higher physiological stress than PRP. Although this was not what we expected, we argue that these findings have important implications.

One potential reason for this pattern is the presence of a minimum performance threshold in order to receive a payment under the fixed payment condition. A minimum threshold to earn the fixed payment has not always been included in previous studies on variable pay (e.g. Dohmen and Falk,

2011). However, we included a threshold to ensure that all participants engaged with the work task to at least a minimal level. All jobs, regardless of whether they use variable or fixed pay, include some kind of performance element, even if it is not explicitly linked to pay as it is in the current study. Unfortunately, structuring the task in this way to ensure that participants have to put some effort into task performance may have caused some stress, which in turn increased the similarity between the two conditions, leading to an effect size that is biased toward finding no effect. Previous studies using the same design (Allan et al., 2021) did see statistically significantly higher levels of both cortisol and self-reported stress in the PRP contracts than in the fixed payment groups, albeit the effects are small. So why do we not see the same pattern here?

The answer is possibly due to an interaction between SET and the minimum performance threshold. It may be that the threshold signifies a clear boundary between “failing” and “passing” and that the risk of being exposed as “failing” to meet this threshold may have inadvertently served as a more potent social-evaluative threat (Fehlner et al., 2020) than being exposed as simply performing at a lower level but without a clear failure threshold as in the fixed payment condition. Indeed, the interaction is primarily driven by the low-performing individuals in our study, as we do not see this pattern in the high-performing group. It may be that the minimum performance threshold acts like a PRP contract, but with a larger incremental step than in the PRP condition.

If both conditions are effectively capturing different types of variable pay and have differing effects depending on the SET context, then it is not accurate to refer to them as low and high-motivated performance conditions. Furthermore, ability plays some part in the direction of the effects that we find. High-performing participants in the PRP condition have almost identical stress coefficients across both the high and low SET conditions, whereas the majority of participants in the fixed payment condition experience a jump in stress when paired with high SET, regardless of their performance. Finally, low-performing PRP participants are somewhat stressed in the control condition yet experience a decrease in stress when their performance is visible to peers. There is some evidence that reducing focus on the individual leads to a lower cortisol response in competitive contexts (Rohleder et al., 2007). It is possible then that the incremental payment reduces focus on the individual in the SET context, but that a clear threshold is perceived as increasing focus in the SET context.

Finally, it is important to note that the interaction effect is not visible in our self-reported stress data. Instead, SET is a significant predictor of higher self-reported stress across both payment conditions, and consequently, the need to differentiate between physiological and perceived stress becomes clear. Prior research (Dickerson et al., 2008) did not always find a correlation between the two, and so it has been suggested that cortisol is particularly driven by self-conscious emotions, such as shame. It may be that incorporating a threshold is not more stressful than variable pay in a traditional sense, but that explicit failure and exposure to others involve more shame. Regardless of how cortisol levels relate to self-reported stress, higher levels of cortisol can have long-term implications for health.

Although the focus of the study was on variable-pay employees, these results still have implications for a large portion of the working population. Many jobs include a quota or a minimum performance threshold in exchange for receiving bonuses or other benefits in the workplace (e.g. in sales; see Schwepker & Good, 2012). Similarly, job promotions may operate in this way—there is a threshold that you have to meet over a period of time to get a promotion. One example is acquiring grant income for academic researchers, a threshold that is often visible to others and that, as suggested by these results, may cause distress if employees are struggling to meet that threshold. Although there is a large literature on stress and demands in the workplace, to date there has been

little research examining the role that minimum thresholds in particular play in causing work-related stress.

This study also has implications for researchers working on PRP from a theoretical perspective. Although we failed to reject the null hypothesis that PRP and SET lead to more stress than other pay contracts, what these results do suggest is that PRP is a much more complex concept, and the differences between various forms of PRP are often not given full consideration in the literature. In the real labor market, PRP includes a wide range of different types of payment contracts. Similarly, the definitions of performance pay in the literature range from piece rate payments only (Davis & Hoyt, 2020) bonuses (Green & Heywood, 2008), as well as consequences for promotion (Baktash et al., 2022), to name a few. We argue that the opposing directions of cortisol reactivity in the different payment contracts depending on whether it is combined with high or low SET show that subtle differences in payment structure can have different effects on physiological stress depending on the context.

Although the experimental design allows us to measure the impact of payment structures on acute stress, this leads to some limitations that need to be discussed. Firstly, the study is restricted to only measuring acute stress, and we cannot draw any conclusions about the impact of payment structures on long-term or chronic stress. Secondly, the study makes use of a student sample, and future research would benefit from expanding the sample beyond university students. Finally, although the math calculation task was chosen as a potentially stressful simulated work task (Dickerson & Kemeny, 2004), the results suggest that anticipation of the task was more stressful than the task itself, in line with the findings by Allan et al. (2021).

In conclusion, the current study aimed to examine whether the combination of a social evaluative threat and PRP payment would increase overall levels of physiological stress. However, in contrast to our original hypothesis, increasing levels of social evaluative threat lead to higher physiological stress in participants completing a work task under fixed payment and not PRP conditions. Importantly, these results do not suggest that stress (physiological or not) is eliminated from variable-payment jobs but rather that various forms of incentivization interact differently with SET. These subtle differences need to be taken into account in future work and stress research, as well as in the workplace.

Acknowledgments

This work was supported by the Economic and Social Research Council under Grant No. [ES/R01163X/1].

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Allan, J. L., Andelic, N., Bender, K. A., Powell, D., Stoffel, S., & Theodossiou, I. (2021). employment contracts and stress: Experimental evidence.

- Journal of Economic Behavior & Organization*, 187, 1–9. <https://doi.org/10.1016/j.jebo.2021.04.015>
- Anđelić, N., Allan, J., Bender, K. A., Powell, D., & Theodossiou, I. (2023). Performance-related pay, mental and physiological health. *Industrial Relations: A Journal of Economy and Society*. Advance online publication. <https://doi.org/10.1111/irel.12334>.
- Aschbacher, K., O'Donovan, A., Wolkowitz, O. M., Dhabhar, F. S., Su, Y., & Epel, E. (2013). Good stress, bad stress and oxidative stress: Insights from anticipatory cortisol reactivity. *Psychoneuroendocrinology*, 38(9), 1698–1708. <https://doi.org/10.1016/j.psyneuen.2013.02.004>
- Baktash, M. B., Heywood, J. S., & Jirjahn, U. (2022). Performance pay and alcohol use in Germany. *Industrial Relations: A Journal of Economy and Society*, 61(4), 353–383. <https://doi.org/10.1111/irel.12301>
- Bain, P., & Taylor, P. (2000). Entrapped by the 'electronic panopticon'? Worker resistance in the call centre. *New Technology, Work and Employment*, 15(1), 2–18. <https://doi.org/10.1111/1468-005X.00061>
- Bender, K. A., & Theodossiou, I. (2014). The unintended consequences of the rat race: The detrimental effects of performance pay on health. *Oxford Economic Papers*, 66(3), 824–847. <https://doi.org/10.1093/oep/gpt032>
- Bosch, J. A., De Geus, E. J. C., Carroll, D., Goedhart, A. D., Anane, L. A., Van Zanten, J. J. V., Helmerhorst, E. J., & Edwards, K. M. (2009). A general enhancement of autonomic and cortisol responses during social evaluative threat. *Psychosomatic Medicine*, 71(8), 877–885. <https://doi.org/10.1097/PSY.0b013e3181baef0519779143>
- Dahl, M. S., & Pierce, L. (2020). Pay-for-performance and employee mental health: Large sample evidence using employee prescription drug usage. *Academy of Management Discoveries*, 6(1), 12–38. <https://doi.org/10.5465/amd.2018.0007>
- Davis, M. E., & Hoyt, E. (2020). A longitudinal study of piece rate and health: Evidence and implications for workers in the US gig economy. *Public Health*, 180, 1–9. <https://doi.org/10.1016/j.puhe.2019.10.021>
- Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *Psychological Bulletin*, 130(3), 355–391. <https://doi.org/10.1037/0033-2909.130.3.355>
- Dickerson, S. S., Mycek, P. J., & Zaldivar, F. (2008). Negative social evaluation, but not mere social presence, elicits cortisol responses to a laboratory stressor task. *Health Psychology*, 27(1), 116–121. <https://doi.org/10.1037/0278-6133.27.1.116>
- Dohmen, T., & Falk, A. (2011). Performance pay and multidimensional sorting: Productivity, preferences, and gender. *American Economic Review*, 101(2), 556–590. <https://doi.org/10.1257/aer.101.2.556>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2007). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/bf03193146>
- Fehlner, P., Bilek, E., Harneit, A., Böhringer, A., Moessnang, C., Meyer-Lindenberg, A., & Tost, H. (2020). Neural responses to social evaluative threat in the absence of negative investigator feedback and provoked performance failures. *Human Brain Mapping*, 41(8), 2092–2103. <https://doi.org/10.1002/hbm.24932>
- Goldberg, D. P., Gater, R., Sartorius, N., Ustun, T. B., Piccinelli, M., Gureje, O., & Rutter, C. (1997). The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychological Medicine*, 27(1), 191–197. <https://doi.org/10.1017/s0033291796004242>
- Goldberg, D., & Williams, P. (1988). *A user's guide to the GHQ*. nFER-Nelson.
- Green, C., & Heywood, J. S. (2008). Does performance pay increase job satisfaction? *Economica*, 75(300), 710–728. <https://doi.org/10.1111/j.1468-0335.2007.00649.x>
- Lazear, E. P. (2000). Performance pay and productivity. *American Economic Review*, 90(5), 1346–1361. <https://doi.org/10.1257/aer.90.5.1346>
- McEwen, B. S. (1998). Stress, adaptation, and disease: Allostasis and allostatic load. *Annals of the New York Academy of Sciences*, 840(1), 33–44. <https://doi.org/10.1111/j.1749-6632.1998.tb09546.x>
- Nicolson, N. (2008). Measurement of cortisol. In L. J. Luecken & L. G. Gallo (Eds.), *Handbook of Physiological Research—Methods in Health Psychology* (pp. 37–73). SAGE Publications.
- Rohleder, N. (2014). Stimulation of systemic low-grade inflammation by psychosocial stress. *Psychosomatic Medicine*, 76(3), 181–189. <https://doi.org/10.1097/psy.0000000000000049>
- Rohleder, N., Beulen, S. E., Chen, E., Wolf, J. M., & Kirschbaum, C. (2007). Stress on the dance floor: The cortisol stress response to social-evaluative threat in competitive ballroom dancers. *Personality & Social Psychology Bulletin*, 33(1), 69–84. <https://doi.org/10.1177/0146167206293986>
- Schwepker, C. H., & Good, D. J. (2012). Sales quotas: Unintended consequences on trust in organization, customer-oriented selling, and sales performance. *Journal of Marketing Theory and Practice*, 20(4), 437–452. <https://doi.org/10.2753/MTP1069-6679200406>