

# Effects of Temporary Mark Withholding on Academic Performance

*Psychology Learning & Teaching*

0(0) 1–15

© The Author(s) 2021



Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/1475725721999958

[journals.sagepub.com/home/plj](https://journals.sagepub.com/home/plj)**Carolina E. Kuepper-Tetzel** 

School of Psychology, University of Glasgow, UK

**Paul L. Gardner**

School of Psychology &amp; Neuroscience, University of St Andrews, UK

## Abstract

Although feedback engagement is important for learning, students often do not engage with provided feedback to inform future assignments. One factor for low feedback uptake is the easy access to grades. Thus, systematically delaying the grade release in favor of providing feedback first—*temporary mark withholding*—may increase students' engagement with feedback. We tested the hypothesis that temporary mark withholding would have positive effects on (a) future academic performance (Experiments 1 and 2) and (b) feedback engagement (Experiment 2) in authentic psychology university settings. For Experiment 1, 116 Year 2 students were randomly assigned to either a Grade-before-feedback or Feedback-before-grade condition for their report in semester 1 and performance was measured on a similar assessment in semester 2. In Experiment 2, a Year 3 student cohort (*t*) was provided with feedback on their lab report before marks were released in semester 1 (mark withholding group,  $N = 97$ ) and compared to the previous Year 3 cohort (*t-1*) where individual feedback and grades were released simultaneously (historical control group,  $N = 90$ ). Using this multi-methodological approach, we reveal positive effects of temporary mark withholding on future academic performance and students' feedback engagement in authentic higher education settings. Practical implications are discussed.

## Keywords

Assessment strategy, mark withholding, use of feedback

---

## Corresponding author:

Carolina E. Kuepper-Tetzel, School of Psychology, University of Glasgow, 62 Hillhead Street, Glasgow G12 8QB, UK.

Email: [carolina.kuepper-tetzel@glasgow.ac.uk](mailto:carolina.kuepper-tetzel@glasgow.ac.uk)

A central component of student learning is to use provided feedback (Hattie & Timperley, 2007). A wide range of research has shown that feedback on previous assessments informs future assessments and enhances academic performance (e.g., Azevedo & Bernard, 1995; Elawar & Corno, 1985; Graham et al., 2015; Lysakowski & Walberg, 1982). While many studies attest to the benefit of feedback for learning, there seems to exist a discrepancy between feedback provision by instructors and feedback uptake by students (Handley et al., 2011; Pitt & Norton, 2017): In many cases, students do not engage with the feedback to inform future assignments—missing out on the opportunity to develop their skills. Although improving feedback provision by making sure the feedback is of high quality is important (O’Neill, 2000), it should be noticed that this is only the first step in the dialogic feedback loop (e.g., Price et al., 2011). The second step is the uptake of the feedback by students. The best feedback is moot if it is not used by students to advance their performance. It has been acknowledged that student feedback uptake is complex and involves cognitive and emotional processes (Evans, 2013; Pitt & Norton, 2017). Winstone et al. (2017) undertook an extensive systematic review of the effective feedback literature published between 1985 and 2014 and identified key features of the receiver (i.e., student), sender (i.e., teacher/lecturer), message, and context of feedback that contribute to student feedback uptake. They found, for instance, that low student feedback engagement can be partly explained by characteristics on the student side such as insufficient feedback literacy (Carless & Boud, 2018) or emotional unreadiness (Evans, 2013). In their proposed model, Winstone et al. (2017) further connect these features to processes in the student (e.g., self-regulation) and feedback interventions (e.g., ways to deliver feedback). The current research explores the triangular interaction between student characteristics, their cognitive processes, and a feedback intervention. More specifically, we focus on the tendency of students to prioritize grades at the expense of processing the written comments and feedback (Jackson & Marks, 2016) and investigate temporary mark withholding—the systematic delaying of marks release in favor of providing feedback and comments first—as a potential feedback intervention.

Why should temporary mark withholding make a difference? Research has shown that an excessive focus on grades can interfere with the students’ ability to self-assess (Taras, 2001)—a crucial cognitive process in the feedback loop. Prioritizing written teacher comments can support students to understand their strengths and weaknesses allowing to allocate effort to aspects that need improvement. This important process can be undermined by seeing a grade. Another issue with focussing on grades at the expense of written feedback is that in case of disappointment—that is, the obtained grade being lower than expected—students may decide not to engage with the written comments at all (Winstone et al., 2017). Consequently, the easy access to grades can decrease feedback engagement. Indeed, Mensink and King (2020) showed a decreased engagement with written feedback if grades can be accessed independently. They used a learning analytics approach and analyzed the student log data recorded in the virtual learning environment to measure student interaction with the feedback across 32 pieces of assessment over 3 undergraduate years across 20 different degrees pathways. A total of 1462 assignments and 484 students were analyzed for this study. For the key analysis on feedback engagement they compared coursework for which marks could be accessed without accessing the written feedback and coursework for which the marks were embedded within the file that held the written feedback: when the mark could be accessed without opening the written comments, students only accessed the feedback 58% of the time. In contrast, written comments were accessed in 83% of

the cases when the marks were embedded within the written feedback. This demonstrates that students prioritize marks, even though their performance in the future would improve more through engagement with written comments (Black & Wiliam, 1998; Page, 1958). In one study, Butler (1988) showed that student performance increased more by presenting written feedback by itself instead of presenting written feedback accompanied with marks or marks alone.

Consequently, temporary mark withholding is one approach that has been suggested as a way to increase students' engagement with feedback. Sendziuk (2010) introduced mark withholding to two cohorts of history students by returning assignments with written feedback only and requiring students to self-assess their assignment based on provided feedback. Students were given the marking criteria, too, and had to write a 100-word justification of their self-assessment. A week later the marks were released and short face-to-face meetings were offered. Sendziuk evaluated his approach via a student questionnaire. He found that 61.6% of the students agreed that withholding marks combined with the feedback engagement task had made them take more notice of the tutors' written feedback. The following quote nicely sums up the mechanisms through which mark withholding can be beneficial:

some noted that they were effectively forced to read the feedback in order to comply with the task (which is not necessarily a bad thing when the learning outcome is so desirable) but others genuinely appreciated the opportunity to engage with the feedback and saw merit in continuing to do so. (Sendziuk, 2010, p. 324)

Jackson and Marks (2016) investigated student feedback engagement and academic performance after introducing written reflections and mark withholding in their postgraduate master-level curriculum in medicine. Feedback engagement was measured via questionnaires and focus groups. The written reflection interventions asked students to write a 400-word piece on the received feedback identifying points for improvements and strengths. One intervention included mark withholding and required students to submit the reflection before receiving the mark. Thus, while a pure contribution of mark withholding cannot be estimated from their study because written reflection was a component in all interventions, the survey data shed light on student attitudes towards mark withholding. Asked about their views on mark withholding, students offered that it made them concentrate better on the written feedback and that a disappointing mark could decrease likelihood to read feedback because of frustration. One student stated: "If I had the grade I would be more focussed on that" (Jackson & Marks, 2016, p. 542). These findings nicely align with the points raised earlier of why temporary mark withholding may be beneficial. Although the majority of students (57%) welcomed mark withholding as an approach, not all student comments were positive towards mark withholding—with critical attitudes surrounding around uncertainty about obtained grades.

While most studies looked at the association between the presence of marks and the feedback processing and academic performance, it is important to investigate if there is a direct effect of marks availability on feedback processing and performance. Lipnevich and Smith (2008) investigated the causal link between these variables in a field experiment with college students enrolled in Introduction to Psychology courses and manipulated the type of feedback and mark provision on a 500-word essay assignment. Students were randomly assigned to no feedback, instructor feedback, or computer-generated feedback condition. Importantly, for the current objective, feedback was provided either with marks or without

marks. In addition, “praise” was added as a variable with students receiving praise or no praise on their work. Unsurprisingly, the findings revealed a strong effect of feedback as student performance increased when feedback was provided. They also found a positive effect of not providing grades: student performance between draft and final version of the essay increased more when grades were not provided. Interestingly, when praise was included, students’ performance was not affected by whether grades were provided or not. However, when no praise was given, students performed better when no grade was provided than when they were provided. This shows that when we deal with a range of instructors who may not add praise to their feedback, withholding grades can be beneficial. Finally, an interaction between type of feedback and grade provision was found: whilst there was no difference between the grade and no grade condition on final performance when no feedback or computer feedback was provided, students performed best after receiving instructor feedback without grades. Thus, Lipnevich and Smith established positive effects of mark withholding on academic performance in an authentic learning setting—where feedback is usually provided by instructors.

## Overview of Experiments

The current research aims at adding to the evidence by investigating the effects of temporary mark withholding on feedback engagement and report writing performance in undergraduate psychology programs. Specifically, we conducted a field experiment and a quasi-experimental field study that tested the hypothesis that mark withholding would have a positive effect on: (a) future academic performance (Experiments 1 and 2); and (b) the engagement with feedback (Experiment 2). The two experiments were run at different UK universities. For the field experiment, a cohort of Year 2 students ( $N = 163$ ) was randomly assigned to one of two feedback conditions for their lab report in semester 1: Grade-before-feedback versus Feedback-before-grade. Performance was measured on a similar assessment in semester 2. For the quasi-experimental field study, a Year 3 student cohort ( $t$ ) was provided with written feedback on an assignment in semester 1 three days before their marks were released ( $N = 102$ ). This cohort was compared to historical data of the Year 3 cohort from the previous year ( $t-1$ ) where individual feedback and grades were released simultaneously ( $N = 95$ ). Feedback consisted of on-script comments and overall comments that elaborated on what could have been improved and what to focus on going forward. Change in performance between the assignments (practical reports) in semesters 1 and 2 was measured and feedback view learning analytics data was analyzed as a proxy for feedback engagement. Data files of both experiments are available on the OSF ([osf.io/axgr4](https://osf.io/axgr4)).

## Experiment 1

### Methods

**Participants.** A cohort of second-year undergraduate psychology students ( $N = 163$ ) from a Scottish university was included in this field experiment.<sup>1</sup> The cohort was randomly assigned to a Grade-before-feedback or Feedback-before-grade group after submission of the students’ semester 1 laboratory report. The same cohort of students was again assessed on the students’ laboratory report for semester 2. Students who failed to submit on time or who

had extensions for their work were excluded from the analysis ( $N=47$ )—resulting in a total sample size of  $N=116$ . Neither participants nor markers were aware of the nature of the study until the conclusion of the study. The research was approved by the ethics committee as a service evaluation of a teaching approach and supported by the Proctor's Teaching Development Award Scheme.

**Material.** The target assignment was a 1500-word laboratory report in semester 1 and semester 2. Students attended two three-hour laboratory classes before submitting their report. The reports required them to answer a core research question, analyze data, and write up a full report consisting of abstract, introduction, methods, results, and discussion. Students worked individually throughout the process and were also required to write the report individually. For the semester 1 report, students collected data using a computer-based version of the Stroop Test and the analysis focused on conducting a mixed design Analysis of Variance. For the semester 2 report, students collected data using a paper-based version of an Autobiographical Memory test and the analysis also focused on conducting a mixed design Analysis of Variance. Reports are marked on a 0–20 scale.

**Design and Procedure.** There were two groups in this field experiment: Grade-before-feedback ( $N=61$ ) versus Feedback-before-grade ( $N=55$ ). Three days prior to the official marks release date, students in each of these groups received their Grade or Feedback via their online Module Management System and were informed that their full Grade/Feedback would be released on the officially agreed date and time three days later. Students submitted their assignments via TurnItIn. Mark withholding was implemented for the report assignment in semester 1 and changes in performance were measured compared to the report assignment in semester 2. For both cohorts, marking was anonymous with one member of staff being responsible for the supervision of four markers from a trained pool of demonstrators who also assist in the delivery of the associated laboratory classes. The feedback to students consisted of individual on-script comments and general comments. Rigorous moderation processes were in place to ensure consistency of marks between and within markers in both cohorts. The markers were not informed of the nature of the study until after completion of the semester 2 marking and moderation process.

## Results Experiment 1

All analyses were run with R in RStudio (RStudio Team, 2019). A significance level of  $\alpha = .05$  was assumed for all analyses and partial eta-squared  $\eta_p^2$  are reported for effect sizes ( $\eta_p^2 = .01$  (small),  $\eta_p^2 = .06$  (medium),  $\eta_p^2 = .14$  (large)) (Cohen, 1988). Bonferroni corrections were applied to all post-hoc pairwise comparisons. Normality of distribution was checked with Q-Q-plots and can be assumed. R analyses were supported with the tidyverse package (Wickham et al., 2019). The afex package (Singmann et al., 2020) was used to run the ANOVA.

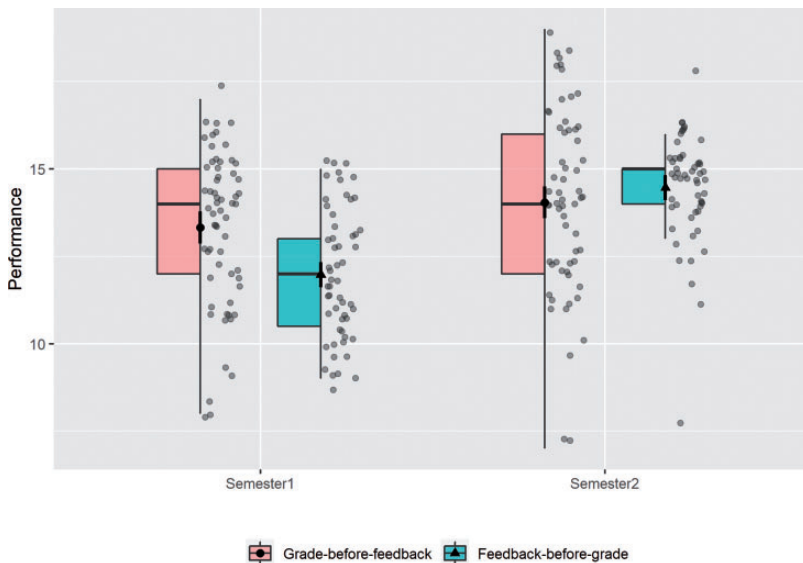
**Effects of Mark Withholding on Report Performance.** The overall performance for the Grade-before-feedback and Feedback-before-grade groups for semester 1 were  $M=13.3$  ( $SE=0.28$ ) and  $M=12.0$  ( $SE=0.25$ ), respectively. The overall performance for semester 2 were  $M=14.0$  ( $SE=0.34$ ) for the Grade-before-feedback group and  $M=14.5$  ( $SE=0.21$ ) for the Feedback-before-grade group (see Table 1 and Figure 1).

A 2 (Condition: Grade-before-feedback vs. Feedback-before-grade) × 2 (Time: semester 1 vs. semester 2) mixed-plot ANOVA showed that there was no main effect of Condition overall,  $F(1,114) = 1.92, p = .169, \eta_p^2 = 0.02$ . There was a significant main effect of Time,  $F(1,114) = 59.69, p < .001, \eta_p^2 = 0.34$ . Overall student performance improved in semester 2,  $M = 14.3 (SE = .20)$ , compared to semester 1,  $M = 12.6 (SE = .20)$ , irrespective of whether grades or feedback was provided first. Most importantly, a significant interaction between Condition x Time was found,  $F(1,114) = 18.11, p < .001, \eta_p^2 = 0.14$ . While there was an increase in performance between semester 1 and semester 2, this increase was steeper in the Feedback-before-grade condition,  $t(114) = 2.52, p = 0.013$ , than in the Grade-before-Feedback condition,  $t(114) = 8.26, p < .0001$ , which supports the proposition that provision of feedback in the absence of a grade had a more positive effect on the degree of performance improvement between assessments.

We would like to point out that despite randomly assigning students to the two different conditions, there was a significant difference in semester 1 report performance between the two groups,  $t(113.7) = 3.59, p < 0.001$ , with students in the Grade-before-feedback group

**Table 1** Report Performance Means and Standard Deviations for the Feedback-Before-Grade and the Grade-Before-Feedback Groups in Semester 1 and Semester 2

Group	Semester	Mean (SD)
Feedback-before-grade	Semester 1	11.96 (1.87)
	Semester 2	14.45 (1.52)
Grade-before-feedback	Semester 1	13.31 (2.18)
	Semester 2	14.03 (2.66)



**Figure 1.** Boxjitter plot of the report performance of students in semester 1 and semester 2 in the two conditions Grade-before-feedback and Feedback-before-grade in Experiment 1.



( $M = 13.3$ ,  $SE = 0.28$ ) performing better in the semester 1 report than students in the Feedback-before-grade group ( $M = 12.0$ ,  $SE = 0.25$ ). Thus, it is possible that students in the Feedback-before-grade group were more motivated to perform better in the semester 2 report because they performed less well than students in the Grade-before-feedback group. We partially tested this alternative explanation by looking at the correlation between semester 1 performance and the improvement between semesters 1 and 2. Indeed, we found negative correlations between semester 1 performance and Sem2-1 improvement for both conditions, Grade-before-feedback,  $r(59) = -.367$ ,  $p = .004$ , and Feedback-before-grade,  $r(53) = -.667$ ,  $p < .001$ . Thus, students who performed less well in the semester 1 report showed more increase in performance in semester 2 and this negative correlation was more pronounced in the Feedback-before-grade condition than in the Grade-before-feedback condition,  $Z = 2.21$ ,  $p = .027$ .

Since we could not completely rule out this alternative explanation for our findings in Experiment 1, we conducted a second experiment using a different methodological approach and also assessed whether students viewed their feedback as an additional measure.

## Experiment 2

### Methods

**Participants.** Two cohorts of third-year undergraduate psychology students ( $N = 197$ ) from a Scottish university were included in this quasi experiment.<sup>2</sup> The intervention cohort (year  $t$ ) consisted of 102 third-year students (mark withholding group) and the control cohort (year  $t-1$ ) consisted of 95 students (historical control group). Students who repeated the third year of their studies and were part of both cohorts were excluded from the analyses ( $N = 5$ ). Students were not made aware of the study, to avoid biased behavior. The research was approved by the ethics committee as evaluation of a teaching approach.

**Material.** The target assignment was a 2000-word report in semester 1 and semester 2. Students attended three two-hour tutorials before submitting their report. The reports required them to develop a research question, analyze data, and write up a full report consisting of introduction, methods, results, and discussion. Students worked in groups during the first two stages but were required to write the report individually. For the semester 1 report, students collected data using a self-perception survey and the analysis focused on conducting an Analysis of Variance. For the semester 2 report, students were given a large dataset containing survey data from a longitudinal, random controlled trial investigating children's cognitive and behavioral abilities. Reports are marked on a 0–23 scale with 23 being the highest grade to be achieved.

**Design and Procedure.** There were two groups in this quasi-experimental design: a historical control group and a mark withholding group. The mark withholding group experienced the intervention in year  $t$  and for the control group we used historical data from the previous year's cohort ( $t-1$ ) where individual feedback and grades were released simultaneously (historical control group). Three days prior to the official marks release date, students in the mark withholding group received an announcement informing them that their individual report feedback had been made available for viewing. They were encouraged to read their feedback and told that their marks would be published in three days' time. Students

submitted their assignments via TurnItIn on Blackboard. Mark withholding was implemented for the report assignment in semester 1 and changes in performance were measured compared to the report assignment in semester 2. For both cohorts, marking was anonymous with a team of two members of staff being responsible for marking of the reports per semester. The feedback to students consisted of individual on-script comments and general comments. In the historical control group, the marking teams in semesters 1 and 2 consisted of four different lecturers. In the mark withholding group, the marking teams in semesters 1 and 2 consisted of three different lecturers—with one marker overlapping in semesters 1 and 2. Rigorous moderation processes were in place to ensure consistency of marks between and within markers in both cohorts. The markers who implemented mark withholding in year  $t$  semester 1 were informed about the general procedure, but no directed hypotheses were discussed. Because of the nature of this quasi-experimental design, we used Year 2 performance of the students in the two cohorts to control for prior academic performance. We accessed learning analytics data of whether students viewed their feedback or not as a proxy for feedback engagement after the semester 2 of the year  $t$  cohort.

### Results Experiment 2

All analyses were run with R in RStudio (RStudio Team, 2019). A significance level of  $\alpha = .05$  was assumed for all analyses and partial eta-squared  $\eta_p^2$  are reported for effect sizes ( $\eta_p^2 = .01$  (small),  $\eta_p^2 = .06$  (medium),  $\eta_p^2 = .14$  (large)) (Cohen, 1988). Bonferroni corrections were applied to all post-hoc pairwise comparisons. Normality of distribution was checked with Q-Q-plots and can be assumed. R analyses were supported with the tidyverse package (Wickham et al., 2019). The afex package (Singmann et al., 2020) was used to run all ANCOVAs. Specifically, for the first analysis, we tested the change in Year 3 report performance in semester 1 to the Year 3 report performance in semester 2 in the historical control group compared to the mark withholding group. We included the previous Year 2 overall performance as a covariate. For the second analysis, we tested the difference in feedback views in the two groups as a proxy for feedback engagement—again controlling for prior academic performance in Year 2.

**Effects of Mark Withholding on Report Performance.** The unadjusted means and standard deviations for report performance in both groups are presented in Table 2. A 2 (Condition: historical control vs. mark withholding)  $\times$  2 (Time: semester 1 vs. semester 2) mixed-plot ANCOVA with Year 2 performance as covariate revealed a significant main effect of Condition,  $F(1,170) = 4.04$ ,  $p = .046$ ,  $\eta_p^2 = .023$ , with students in the historical control cohort performing slightly worse ( $M = 15.5$ ,  $SE = 0.21$ ) than students in the mark withholding group ( $M = 16.1$ ,  $SE = 0.21$ ). There was a significant main effect of semester,  $F(1,170) =$

**Table 2** Report Performance Means (Unadjusted) and Standard Deviations for the Historical Control and the Mark Withholding Groups in Semester 1 and Semester 2

Group	Semester	Mean (SD)
Historical control	Semester 1	15.18 (2.33)
	Semester 2	15.28 (2.91)
Mark withholding	Semester 1	15.59 (2.77)
	Semester 2	17.01 (2.52)

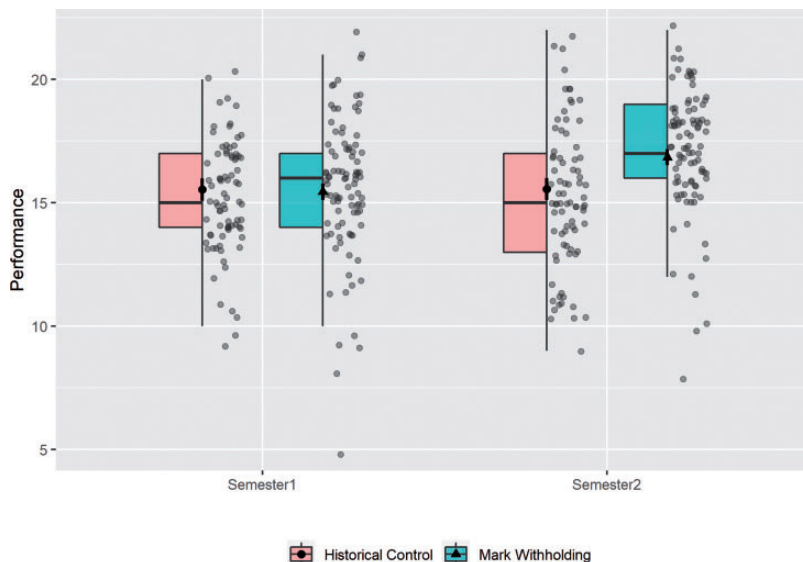


13.16,  $p < .001$ ,  $\eta_p^2 = .072$ , with students performing better in the semester 2 report ( $M = 16.2$ ,  $SE = 0.18$ ) than in semester 1 report ( $M = 15.5$ ,  $SE = 0.18$ ). Unsurprisingly, Year 2 performance significantly predicted report performance in Year 3,  $F(1,170) = 76.93$ ,  $p < .001$ ,  $\eta_p^2 = .312$ . Most importantly, however, there was a significant interaction between Condition and Semester,  $F(1,170) = 11.95$ ,  $p < .001$ ,  $\eta_p^2 = .066$ . The interaction is visualized in Figure 2. In line with the prediction, there was a significant increase in report performance between semester 1 ( $M = 15.4$ ,  $SE = 0.25$ ) and semester 2 ( $M = 16.8$ ,  $SE = 0.25$ ) in the mark withholding group,  $t(170) = 5.22$ ,  $p < .0001$ . However, in the historical control group, report performance was similar in both semesters,  $t(170) = 0.08$ ,  $p = .935$ .

**Effects of Mark Withholding on Feedback Views.** We accessed whether students viewed their feedback of the semester 1 report by obtaining the feedback view data from Blackboard. A one-way ANCOVA with Condition as independent variable and Year 2 performance as covariate on proportion of students who viewed their feedback revealed a significant main effect of Condition,  $F(1,182) = 12.76$ ,  $p < .001$ ,  $\eta_p^2 = .066$ , with more students viewing their feedback in the mark withholding group ( $M = 95\%$ ,  $SE = 3.32$ ) than in the historical control group ( $M = 78\%$ ,  $SE = 3.48$ ). There was no effect of the covariate on feedback views,  $F(1,182) = 2.01$ ,  $p = .16$ ,  $\eta_p^2 = .011$ .

**Mediation of Effects of Mark Withholding on Report Performance via Feedback Views.** We ran a mediation analysis to test whether the effects of mark withholding on report performance would be partially mediated by whether students viewed their feedback. We used the mediation R package (Tingley et al., 2014) to run this analysis.

We found that the average direct effect of mark withholding on report performance was still significant,  $ADE = 1.28$ ,  $p = .004$ , and that the average causal mediation effect of

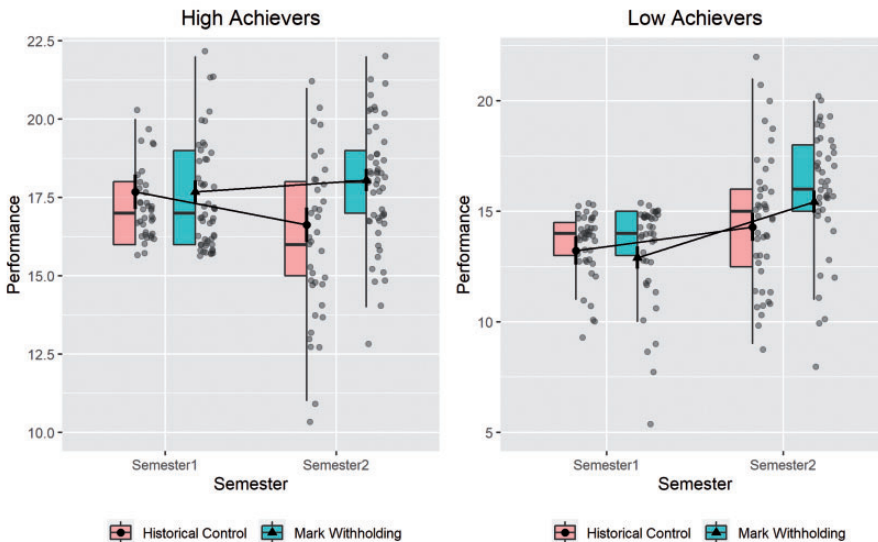


**Figure 2.** Boxjitter plot of the report performance of students in semester 1 and semester 2 in the two groups, Historical control and Mark withholding, in Experiment 2.

feedback views was not significant,  $ACME = 0.096$ ,  $p = .364$ . The proportion mediated via feedback views in the model was only 7%.

**Exploratory Analyses.** We explored whether the interaction revealed for the report performance holds for students who obtained grades of 16 and higher (high achievers) (grades As and Bs) versus students who got grades of 15 (grade C) and lower (low achievers) in their semester 1 report.

We found significant interactions between Condition and Semester for both student achievement groups,  $F(1,85) = 10.33$ ,  $p = .002$ ,  $\eta_p^2 = .11$  (high achievers) and  $F(1,82) = 6.86$ ,  $p = .010$ ,  $\eta_p^2 = .077$  (low achievers). However, the nature of the interactions looked quite different for the two student achievement groups (see Figure 3). For the high achievers in the mark withholding group, there was no significant change in performance between semester 1 and semester 2,  $t(85) = 1.31$ ,  $p = .193$ , but for high achievers in the historical control group marks significantly decreased between semester 1 and semester 2,  $t(85) = -3.15$ ,  $p = .002$ . Low achievers, on the other hand, increased their performance between semester 1 and semester 2 in both, the historical control,  $t(82) = 2.74$ ,  $p = .008$ , and mark withholding group,  $t(82) = 6.40$ ,  $p < .001$ , but the increase was more pronounced in the mark withholding group. Furthermore, looking at feedback views in the two student achievement groups, more students viewed their feedback in the mark withholding group than in the historical control group, but the difference was only significant for the high achievers ( $M_{MarkWithholding} = 100\%$  vs.  $M_{HistoricalControl} = 81\%$ ),  $F(1,92) = 12.09$ ,  $p < .001$ ,  $\eta_p^2 = .116$ , not for the low achievers ( $M_{MarkWithholding} = 90\%$  vs.  $M_{HistoricalControl} = 75\%$ ),  $F(1,87) = 3.42$ ,  $p = .068$ ,  $\eta_p^2 = .038$ .



**Figure 3.** Boxjitter plots of the report performance of students in semester 1 and semester 2 in the two groups, Historical control and Mark withholding, split by student achievement level. High Achievers are students who achieved a grade of B or higher and Low Achievers achieved a grade of C or lower in their semester 1 report in Experiment 2.

## Discussion

The results are in line with the hypothesis and reveal that students who received their feedback before their grades showed a greater gain in performance between semesters 1 and 2 compared to the students who received their grades before the feedback.

Experiment 1 was a fully-fledged field experiment where students of the same cohort were randomly assigned to either Feedback-before-grade or Grade-before-feedback conditions. We show that presenting feedback prior to the grade led to an improvement in performance in the following assessment. Two explanations were suggested for this effect: first, this may have been due to students wishing to self-assess (i.e., estimate their grade) their performance by looking at feedback when they have not been given their grade and this engagement with their feedback may have led to an increase in performance in the following semester. However, we found partial evidence that it could also be the case that because the Feedback-before-grade group had lower performance overall in first semester than the Grade-before-feedback group, individuals may have been motivated to work harder to improve their grades. Unfortunately, the current data from Experiment 1 does not allow us to disentangle both explanations and provide a definite answer. However, in Experiment 2, we took a different methodological approach and also analyzed students' feedback engagement by analyzing learning analytics data to see whether temporary mark withholding influenced students' feedback viewing behavior.

For Experiment 2 we conducted a quasi-experimental field study and analyzed feedback viewing and student performance data. Controlling for previous academic performance, we found that the mark withholding group showed an increase in performance between the two report assignments in semesters 1 and 2 whereas the historical control group did not show such an increase. In the mark withholding group this translates into an average increase of 1.4 points or 6% on a 23-point grade scale; moving students on average from a C1 to almost a B2. In addition, and in line with the expectations, the mark withholding group showed a higher proportion of feedback views than the historical control group. We followed this up with a mediation analysis and found no significant mediation effect of feedback views on the effect of mark withholding on change in performance. At first this seems to rule out our explanation that the underlying process for the benefits of mark withholding is that students will engage more with the feedback, which, in turn, enhances future performance. However, we would refrain from drawing this conclusion based on our experiment alone: we used a simple proxy for feedback engagement in our study by just counting whether students viewed the feedback or not. We have no measure of what students did with the feedback and how they engaged with it. Thus, while it is interesting to see that more students viewed their feedback in the mark withholding group than in the historical control, this variable is not fine-grained enough to assess qualitative changes in feedback engagement.

We conducted further exploratory analysis and re-ran the analyses separately for students who achieved a mark of B or higher (high achievers) and for students who received a mark of C and lower (low achievers) on the semester 1 report. We found that high and low achievers were affected differently by temporary mark withholding. High achievers showed a decrease in performance between semesters 1 and 2 in the historical control, but not in the mark withholding group. Low achievers showed an increase in report performance from semester 1 to semester 2 in both groups, but the increase was steeper in the mark withholding group than in the historical control group. While both, low and high achievers, showed more feedback views in the mark withholding group than in the historical

control group, this effect was only significant for the high achievers. It is possible that temporary mark withholding increased their engagement with the feedback which supported maintaining their high performance in semester 2—whereas the lower engagement with the feedback in the historical control group was detrimental for the performance in semester 2.

Using different methodological approaches, we find medium-sized effects for benefits of temporary mark withholding in authentic higher education settings—particularly, for psychology report writing. This expands on previous studies showing the benefits of temporary mark withholding (e.g., Butler, 1988; Lipnevich & Smith, 2008). In our experiments, we used the simplest implementation of this approach; that is, temporary mark withholding without any accompanying tasks. On the basis of previous research, however, that highlights the positive outcomes of guiding students through the feedback engagement phase, we would suggest incorporating clear guidance and activities in the phase between written feedback and mark release that support feedback literacy (Carless & Boud, 2018). Moreover, because students tend to focus on grades and want to know their grades, temporary mark withholding could be met with skepticism (Smith & Gorard, 2005). However, openly discussing the rationale behind this approach, managing student expectations, and designing short activities (e.g., in the form of written reflection statements or answering questions about the obtained feedback) can enhance student buy-in into temporary mark withholding as a valuable approach (see Jackson & Marks, 2016; Sendziuk, 2010). A blog post by Loudon (2017) outlines a plan on how to implement temporary mark withholding and specifically points to the importance of giving agency to students during the feedback phase by providing them with opportunities to respond to written feedback before receiving their marks. It would be interesting to investigate these more elaborate feedback engagement ideas in combination with temporary mark withholding in future research. This approach would also result in a more in-depth and elaborated feedback engagement measure than what we used in Experiment 2 as a proxy for feedback engagement (i.e., feedback view learning analytics data that was readily available through the virtual learning environment). Future research should assess the quality of feedback engagement to shed light on its potentially mediating effect of mark withholding on performance.

In conclusion, our findings show positive effects of temporary mark withholding on report performance in psychology for Year 2 and Year 3 students. This converging finding from two experiments is particularly compelling because it was: (a) revealed in an authentic educational setting that captured student learning as it unfolds; and (b) obtained by using different methodological approaches in each of the experiments. Triangulating the findings across experiments using different designs, which each come with their own strengths and weaknesses, increases our confidence in drawing conclusions for teaching practice. We further demonstrate that more students tend to view the written feedback when temporary mark withholding is in place than when marks and written feedback are released together. Future research should investigate the best feedback engagement activity for students to engage between the release of written feedback and the marks. Co-designing this process with students may increase students' appreciation of mark withholding as a useful approach and lead to further increases in feedback engagement. In addition, more research into the potential benefits of mark withholding on academic skills other than report writing should be explored.

In the context of this special issue and in conjunction with the previously reviewed literature, we provide accumulative evidence that temporary mark withholding—a potentially underused practice—can be an effective strategy to improve future student performance and

potentially foster self-regulated learning in university settings. As a practical recommendation, we would endorse temporary mark withholding as a low-cost teaching practice that can be easily implemented by instructors.

### Acknowledgments

The authors would like to thank Dr Heather Branigan and Dr Joshua March for helping to implement the research project during their time as lecturers at the University of Dundee, UK. The authors would further acknowledge that part of this research was supported by the Proctor's Teaching Development Fund at the University of St Andrews, UK.

### Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The authors would further acknowledge that part of this research was supported by the Proctor's Teaching Development Fund at the University of St Andrews, UK.

### ORCID iD

Carolina E Kuepper-Tetzel  <https://orcid.org/0000-0003-0830-7915>

### Notes

1. The total sample size was determined by the number of students in the cohort.
2. The total sample size was determined by the number of students in the cohorts.

### References

- Azevedo, R., & Bernard, R. M. (1995). A meta-analysis of the effects of feedback in computer-based instruction. *Journal of Educational Computing Research*, 13(2), 111–127.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74.
- Butler, R. (1988). Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology*, 58(1), 1–14.
- Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315–1325.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Psychology Press.
- Elawar, M. C., & Corno, L. (1985). A factorial experiment in teachers' written feedback on student homework: Changing teacher behavior a little rather than a lot. *Journal of Educational Psychology*, 77(2), 162–173. <https://doi.org/10.1037/0022-0663.77.2.162>
- Evans, C. (2013). Making sense of assessment feedback in higher education. *Review of Educational Research*, 83(1), 70–120.
- Graham, S., Hebert, M., & Harris, K. R. (2015). Formative assessment and writing: A meta-analysis. *Elementary School Journal*, 115(4), 523–547.
- Handley, K., Price, M., & Millar, J. (2011). Beyond “doing time”: Investigating the concept of student engagement with feedback. *Oxford Review of Education*, 37, 543–560.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81–112.
- Jackson, M., & Marks, L. (2016). Improving the effectiveness of feedback by use of assessed reflections and withholding of grades. *Assessment & Evaluation in Higher Education*, 41, 532–547.

- Lipnevich, A. A., & Smith, J. K. (2008). Response to assessment feedback: The effects of grades, praise, and source of information. *ETS Research Report Series*, 2008(1), 1–57. <https://doi.org/10.1002/j.2333-8504.2008.tb02116.x>
- Louden, K. (2017, June 4). Delaying the grade: How to get students to read feedback. *Cult of Pedagogy*. <https://www.cultofpedagogy.com/delayed-grade/>
- Lysakowski, R. S., & Walberg, H. J. (1982). Instructional effects of cues, participation, and corrective feedback: A quantitative synthesis. *American Educational Research Journal*, 19(4), 559–572.
- Mensink, P. J., & King, K. (2020). Student access of online feedback is modified by the availability of assessment marks, gender and academic performance. *British Journal of Educational Technology*, 51(1), 10–22.
- O'Neill, J. (2000). SMART goals, SMART schools. *Educational Leadership*, 57, 46–50.
- Page, E. B. (1958). Teacher comments and student performance: A seventy-four classroom experiment in school motivation. *Journal of Educational Psychology*, 49(4), 173. <https://doi.org/10.1037/h0041940>
- Pitt, E., & Norton, L. (2017). “Now that’s the feedback I want!” Students’ reactions to feedback on graded work and what they do with it. *Assessment & Evaluation in Higher Education*, 42(4), 499–516.
- Price, M., Handley, K., & Millar, J. (2011). Feedback: Focusing attention on engagement. *Studies in Higher Education*, 36, 879–896. <https://doi.org/10.1080/03075079.2010.483513>
- RStudio Team. (2019). RStudio: Integrated development for R. RStudio, Inc. <http://www.rstudio.com/>
- Sendziuk, P. (2010). Sink or swim? Improving student learning through feedback and self-assessment. *International Journal of Teaching and Learning in Higher Education*, 22(3), 320–330.
- Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2020). afex: Analysis of factorial experiments. R package version 0.27-2. <https://CRAN.R-project.org/package=afex>
- Smith, E., & Gorard, S. (2005). “They don’t give us our marks”: The role of formative feedback in student progress. *Assessment in Education: Principles, Policy & Practice*, 12(1), 21–38.
- Taras, M. (2001). The use of tutor feedback and student self-assessment in summative assessment tasks: Towards transparency for students and for tutors. *Assessment & Evaluation in Higher Education*, 26, 605–614.
- Tingley D., Yamamoto T., Hirose K., Keele L., & Imai K. (2014). mediation: R Package for cCausal mMediation aAnalysis. *Journal of Statistical Software*, 59(5), 1–38. <http://www.jstatsoft.org/v59/i05/>
- Wickham, H., Averick, M., Bryan, J., Chang, W., D’Agostino McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., . . . Yutani, H., et al. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2017). Supporting learners’ agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. *Educational Psychologist*, 52, 17–37.

## Biographical Information and Research Interests

**Carolina Kuepper-Tetzl** is a lecturer in the School of Psychology at the University of Glasgow, UK. She is an expert in applying findings from cognitive science to education and an enthusiastic science communicator. She obtained her PhD in cognitive psychology from the University of Mannheim and pursued postdoctoral positions at York University in Toronto and the Center for Integrative Research in Cognition, Learning, and Education (CIRCLE) at Washington University in St Louis. She was a lecturer in psychology at the University of Dundee for four years before joining the School of Psychology at the



University of Glasgow in January 2020. Her expertise focuses on learning and memory phenomena that allow implementation to educational settings to offer teachers and students a wide range of strategies that promote long-term retention. She is convinced that psychological research should serve the public and, to that end, engages heavily in scholarly outreach and science communication. She is a member of the Learning Scientists and founded the Teaching Innovation & Learning Enhancement (TILE) network. She is frequently invited to give continuing professional development workshops and keynotes on learning and teaching worldwide. She was awarded Senior Fellow of the Higher Education Academy (HEA). She is passionate about teaching and aims at providing her students with the best learning experience possible. Carolina teaches research methods and cognition and promotes service learning in higher education.

**Paul Gardner** is a senior lecturer in the School of Psychology and Neuroscience at the University of St Andrews, UK. He has led on a significant number of projects in applied settings in education and health. He has regularly contributed to the British Psychological Society Flagship Lectures aimed at pre-university students as well as performing at the Edinburgh Fringe Festival. He also runs the Royal Society of Edinburgh-sponsored Scottish-wide creative writing essay competition for high school pupils, 'Science Fiction: Make Believe'. He is a Fellow of the HEA. His main responsibilities are for outreach and science communication and he leads his school's Lifelong Learning Programme as well as their input to the widening access projects. He teaches quantitative methods for the social sciences and has broad interests in mastery learning and skill acquisition.