

Pre-Analysis Plan for: Good or Bad News First? The Effect of Feedback Order on Performance and Motivation

Lavinia Kinne^a

February 12, 2022

This project uses a field experiment to study whether the order of feedback matters for motivation and performance of university students. It relies on a randomized assignment of the order of a positive and a negative feedback element. This way, a causal effect of the ordering on the respective outcomes performance and motivation can be obtained.

1. Study Design

This field experiment will be conducted online with undergraduate (i.e. Bachelor) Economics and Business students at LMU Munich who can voluntarily sign up for additional exam preparation of a specific course.¹ Depending on the course language, all information and questionnaires will be presented either in English or in German.

The following flow chart shows the overall sequence of the study design. The following subsections explain each of the subparts in detail.

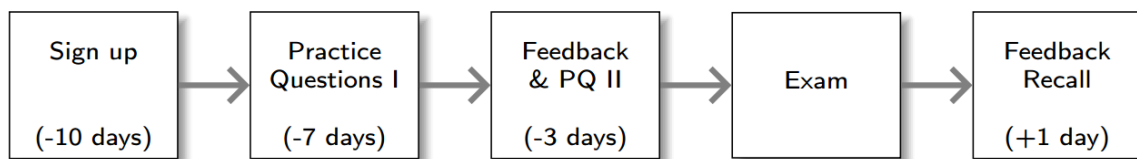


Figure 1: Overview of the study design

¹Currently, two courses have agreed to participate in the Winter Semester 2021/2022. Further courses might be added in the Summer Semester 2022 as described below.

1.1. Sign up

The opportunity to participate in the study will be announced in the lecture as well as via the course's Moodle page. The announcement will contain of the link to the registration form as well as some basic information about the study: students will be informed that they have the opportunity to practice their knowledge of course content in the form of exam-type questions and that they will receive personalized feedback on these questions. Furthermore, they are presented with the chance to win a €50 voucher in case of full participation to all three questionnaires described below. Sign up will then happen via a standardized form in which students state their university email address and give consent to be contacted via the provided email address for the purposes of this research project. Participants will have to use their login-protected university email address in order to guarantee a personalized access. Students can register for the experiment until 10 days before the respective exam.

One day after the deadline for signing up, students receive an email explaining the procedure of the exam preparation. This email will again announce the dates for the additional exam preparation questions (7 days before the exam and 3 days before the exam), of the feedback I will provide (3 days before the exam) as well as the date for a short post-exam questionnaire (one day after the exam). It will be specified that only the participation in all these three elements implies full participation in the experiment and hence leads to the chance of winning a voucher. Each voucher will have a value of €50 and can be chosen from the platforms Amazon, Avocadostore, and Netflix. The number of vouchers will be determined after the final number of participants is settled such that participants have a chance of at least 5% of winning. Finally, students will be informed that they will be recontacted once they have received the grades for the respective exam and that each student who reports the grade together with a proof such as a screenshot will receive an additional €10 voucher.

1.2. First set of practice questions

Seven days before the exam, students will receive an email with a link to the first questionnaire including the first set of exam preparation questions. On the first page, there will be an information text with instructions. These will include both information about data protection and security as well as a reminder of the structure of the study and the chance to win a voucher when completing all parts of the research project. Finally, respondents agree to be recontacted for the next set of questions and to have their data stored and processed according to data protection regulations. Prior to the practice questions, the students will have to answer a series of questions on some general characteristics as well as personality traits. The former will contain questions on age, gender, semester, field of study, their high school GPA and last high-school math grade as a proxy for ability, their parents' highest education and employment status as a proxy for socio-economic status as well as a question on whether students and their parents were born in Germany as a proxy for culture. The surveyed personality traits include patience, risk-aversion, the big 5 personality traits, self-efficacy, and feedback aversion, and will later be used to understand the mechanism of a potential causal effect (see section on Hypotheses and Analyses). Furthermore, students will have to state how motivated they are to study for the exam and in general, their study plan (hours dedicated to studying for this course on each of the six remaining days before the exam as well as so far, per week, and topics on which they want to focus on in their remaining study

time), and which performance they expect for the course (expected grade and relative performance in quartiles of the distribution) and the practice questions (total points as well as relative performance), including a self-evaluation of this expected performance for both. This part of the questionnaire is expected to take not more than 10 minutes. Afterwards, they will be presented with 5 questions about exam-relevant content. These will be adjusted for the current situation of online exams and will allow for the same conditions as the later exam (e.g. open book and formula sheets). Students have 30 minutes to answer the practice questions, a timer will ensure that students stick to this time frame.

The practice questions will then be corrected. A general score (out of 20 on a full-point scale) as well as a ranking of students will be calculated. Furthermore, questions refer to specific topics of the respective course such that each student will also have a sub-score for each of the topics. According to these sub-scores, a within-student ranking of topics can be created, resulting in a topic students responded to best as well as worst.² In the later feedback round, each student in one of the treatment groups will then be informed about her best and worst topic according to this within-person ranking, in varying order. Respondents in the control group will not receive any feedback at this stage, but only after the second set of practice questions together with everyone else. This is to ensure that the experiment can be advertised as a chance to further practice exam knowledge, including personalized feedback.

1.3. Feedback and second set of practice questions

The core part of the experiment is the feedback treatment that will follow four days after the practice questions and three days before the respective exam. Students will receive a (personalized) link to visit a page with an individualized structure. They will either be informed about their best or worst topic in the first set of practice questions first and afterwards about the other one, or vice versa.

The communication of feedback to the participants will look as follows. All students will see a welcoming page, reminding individuals that they can win €50 vouchers when completing all parts of the survey and that in this questionnaire they will receive further practice questions as well as feedback. Participants from the control group will then be directed to questions about beliefs, self-evaluation, and motivation before moving on to the second set of practice questions. Participants from the treatment group will first see their feedback instead. The feedback about a positive and a negative dimension will be split onto separate screens. It is presented in the following way. Students will be informed that their performance on the first set of practice questions has been evaluated and that a total score as well as subscores for each block of questions for each topic have been calculated. On the next page, they will then see the first feedback. Positive (negative) feedback will be communicated as the best (worst) of the topics evaluated on their first set of practice questions. The text will be "According to all topics evaluated on your performance in the first set of practice questions, you did best/worst on XXX", where XXX stands for their personal best/worst out of the respective topics. For the positive part, an additional sentence is added: "This is your personal strength, great!". Similarly,

²In case of equality of scores for the best and/or the worst topic, one of them will be chosen randomly, depending on whether students put in any text at all. For students who leave all questions blank, it might happen that the best and the worst dimension coincide. In case this happens, randomization of best and worst topic for feedback will be repeated until the topics are different.

for negative feedback: "This is your personal weakness, what a pity!". Furthermore, a green (red) thumbs up (down) will be added to make the positive and negative nature of feedback more salient. Examples of how positive and negative feedback elements will look like can be seen in the appendix (in German).

Between the two pieces of feedback as well as after the two feedback parts they will be asked about their feelings regarding the feedback, their beliefs about performance (including a self-evaluation), their current motivation to study for the exam, and their study plan (hours dedicated to studying for this exam on each of the two remaining days before the exam and topics on which to focus). Finally, students will receive an additional set of practice questions which will contain 10 brief multiple choice questions related to exam content. These will have a time limit of 60 seconds each such as to be more similar to the later exam situation. Lastly, all students will be automatically informed about their score in the second set of multiple choice questions (total number of correct answers) right after responding to them. This part of the study should not take more than 30 minutes.

1.4. *Post-exam questionnaire*

One day after the final exam, students will receive a link to a post-exam survey. It will contain questions about their performance beliefs (including a self-evaluation) of the exam, their actual implemented study plan (hours dedicated to studying for this exam on each of the six days before the exam and the topics on which they focused), their perception of the difficulty of the practice questions provided in the experiment as compared to the exam as well as their usefulness as an exam preparation, a recall of the feedback, and how they felt about the feedback in general, including its usefulness. This survey should not be longer than 15 minutes.

As soon as official grades are released, students will again be contacted. They will be asked to report their grade in the respective course together with some sort of proof, e.g. a screenshot of their transcript. Each student will then receive a €10 voucher after a quick check of their grade proof. Grade proofs will be immediately deleted after checking their validity.

1.5. *Treatment groups*

Since the main focus of this study is the effect of feedback ordering on performance and motivation, the control group receiving no feedback on the first occasion will be comparatively small. Its main purpose is to observe time trends such as study progress that are independent of feedback as well as to replicate results from the literature showing the effect of feedback as compared to no feedback. Furthermore, it will be used to analyze reactions to the first feedback elements in the treatment groups. The control group *No feedback* will hence comprise of 20% of the sample or at least 50 individuals in case 20% are less than 50. The remaining participants will be equally distributed between the treatment groups *Feedback positive-negative (POSNEG)* and *Feedback negative-positive (NEGPOS)*. Randomization on an individual level (by course) will be performed after receiving the answers to the first set of practice questions, via STATA software. Also, randomization will be stratified by gender since this dimension is the main interest of this study in terms of heterogeneity.

If a very large number of participants signs up (more than 750), two further treatment groups will be added. Students will receive exactly the same questions as all other

students with one exception: they will receive only one piece of feedback about their best *or* worst dimension whereby feedback will be randomly positive or negative. They will in sum have the same size as the *No feedback* group, i.e. will comprise of 20% of the sample, equally split between positive and negative feedback. The two treatment groups will then comprise of 30% of the sample each. The one-feedback treatment groups allow for comparison of receiving one feedback element as compared to two elements in general and in their varying order.

In case the number of participants remains below 500 in this first try in February 2022, further funds might be used to recruit additional participants in courses conducted in the summer semester of 2022 (exam session in July and August 2022).

1.6. *Exclusion criteria*

Participants will be excluded from the analysis in the following cases. Firstly, students are excluded if they do not consent to being contacted via the provided email address in the sign-up form. Furthermore, students will be excluded if they do not participate in the first and/or the second set of practice questions. Participants who do not respond to the third survey following the exam will be counted as partial participants and might be considered for outcomes and mechanisms surveyed before the third questionnaire. This holds similarly for those who do not report their final grade.

In the sign-up form students have to communicate their university email address, the form does not report an error though in case they use a different email address. For reasons of data protection it is important that participants use their university email address, i.e. to ensure via login to this account that only the students themselves answer the questions. Since respondents give their consent to contact them via the provided email address, those who provided a different email address will be recontacted and asked to provide their university email address. They are finally excluded if they fail to do so after being recontacted. In case a respondent provides an email address with an erroneous domain, a minor correction for obvious typos will be performed as long as they are in the domain of the email and not in the personal information in the first part. Also, duplicates are dropped in case individuals sign up more than once.

Some of the questions about personality traits rely on several sub-questions with reversed coding. These lend themselves to detect inconsistencies in respondents' behavior as well as to check their attention. Respondents with inconsistencies will not be excluded from any analysis, but they will be excluded for a robustness check if they show such an inconsistency in all of the questions relying on reversed coding (see below).

Some questions force participants to provide an answer before being able to move on to the next part of the questionnaire (e.g. gender as well as the consent questions). Other questions instead, might be skipped by respondents such that they will be excluded of all specifications that include the respective characteristics as controls. Furthermore, students who leave empty all of the first 5 practice questions will be excluded from the analysis in a robustness check since for them the feedback is hardly credible, especially the positive part. Similarly, this holds for students who achieve full points in the practice questions.

1.7. *Power Analyses*

The power analyses to determine important sample sizes and minimum detectable effect sizes will be explained in the following paragraph. The underlying calculations have been

performed with the STATA command *power* as well as with Optimal Design Software³. The two courses targeted in the winter semester of 2021/2022 comprise of approximately 800 and 700 students respectively.⁴ The sample size could be increased by using other courses in the following semester (summer semester 2022) which is to be determined depending on sign up in the first session (see above).

This experiment has both between-subject and within-subject elements. Between-subject comparisons aim at understanding the impact of feedback and feedback order on post-feedback outcomes. They rely on comparing group averages of the main outcomes performance and motivation as well as secondary outcomes beliefs (including self evaluation) and study behavior. Since the number of participants in this study cannot be predicted before actually implementing it, power calculations here will refer mostly to possible scenarios and potential cutoffs for further data collection.

As mentioned above, the control group should at least comprise of 50 individuals or 20% of the sample. In case these coincide, the full sample size would be 250, with 100 students in each of the treatment groups. When looking at typical educational datasets such as the international PISA data, one can see that usually individual covariates explain about 30% of the variance in math scores (see e.g. [Hanushek et al. \(2022\)](#)). Considering this for the design of my experiment and assuming a significance level of 5% as well as a power of 80%, this would imply a minimum detectable effect size (MDES) of around 0.35 standard deviations (sds). This is indeed a rather large effect size, but similar effect sizes have been found in papers with feedback interventions, especially on outcomes such as motivation. For example, [Azmat et al. \(2019\)](#) find an effect of one third of a sd on satisfaction, their treatment similarly only varies the type of feedback rather than giving or not giving feedback at all. Similarly, [Hermes et al. \(2021\)](#) find an effect of 0.47 sds of relative performance feedback on motivation for low-achievers in a setting with primary school students. The cutoff mentioned for further data collection in summer 2022 refers to 500 students, i.e. 200 in each of the treatment groups. This corresponds to a MDES of 0.24. Similarly, a full sample size of 750 students (300 in each treatment group) implies a MDES of 0.19. This effect size compares to the effects of feedback on performance for university grades in a rather uncontrolled natural experiment setting of 0.13 sds in [Bandiera et al. \(2015\)](#) as well as those of a more controlled but more long-term and low-intensity treatment of 0.2 sds in [Azmat et al. \(2019\)](#).

Since all outcome measures are also elicited before treatment, a within-individual comparison is possible as well. Motivation, beliefs, and planned study behavior are elicited before the individuals perform the first set of practice questions and can hence be used as a pre-treatment performance measure. All pre-treatment outcome measures can both be used as control variables in the main treatment regressions as well as for a within-person design where deviations from prior outcome values can be detected. Both of these procedures should increase the power of the presented inference.

Finally, attrition is most likely going to be a problem in this context, even though participation is to individuals' personal benefit on many levels. To prevent attrition, especially in the later parts without practice questions, students will be separately incentivized for reporting their grades and will receive reminders for all questionnaires.

³Raudenbush, S. W., et al. (2011). Optimal Design Software for Multi-level and Longitudinal Research (Version 3.01) [Software]. Available from www.wtgrantfoundation.org

⁴A few students from a relatively small major are supposed to take both courses in the same semester. For those who sign up for the experiment for both courses, only the first observation will be used. Similarly, this will be done for all other students who sign up for both courses for other reasons.

Furthermore, the post-exam questionnaire as well as grade reporting have a longer deadline (around one week).

2. Measurement of Variables

2.1. Outcome variables

The primary outcomes of this study are performance and motivation, secondary outcomes consist of beliefs (including self evaluation) and study behavior. The following list explains how these are measured.

1. Motivation:

- "How motivated are you to study for the exam of this course [name of course]?" (not motivated at all/rather not motivated/neutral/rather motivated/very motivated)
- "How motivated are you to study in general for exams of your university studies?" (not motivated at all/rather not motivated/neutral/rather motivated/very motivated)

The first question will be posed multiple times to the students: before answering the first set of practice questions as well as after each of the feedback elements. The second question will only be asked before the first set of practice questions.

2. Performance:

- total score of students on the first set of practice questions (out of 20, pre-feedback)
- total score of students on the second set of practice questions (out of 10, post-feedback)
- grade in final exam (German scale, post-feedback)

3. Beliefs:

- expected total score on the first set of practice questions (out of 20)
- expected total score on the first set of practice questions (out of 10)
- expected final exam grade (German scale)
- for all three: expected quartile of the later performance distribution (probabilities have to add up to 100)
- for all three: self evaluation of expected performance (*very poor*, *poor*, *neutral*, *good*, or *very good* as in [Exley and Kessler \(2022\)](#))

Beliefs (and self-evaluation) for the first set of practice questions are asked before and after solving the questions as well as after each feedback element. The same holds true for beliefs about exam performance. For the second set of practice questions they are asked right before solving it, i.e. after both feedback elements have been provided.

4. Study plan

- how many hours they have dedicated to studying for this exam so far (per week)
- how many hours they are planning to spend studying in each of the remaining days before the exam
- topics they are planning to focus on in the remaining study time

The first question will only be asked in the first questionnaire, i.e. before treatment. The other two elements are asked both before treatment (7 days before the exam) as well as after each of the feedback elements (3 days before the exam). The hours referred to in the second question will be asked for the respective days remaining before the exam. Furthermore, students are asked again how many hours they actually studied for the exam on the 6 days before the exam in the post-exam questionnaire.

All outcomes besides performance are also measured between the feedback elements. This allows for a subanalysis of these outcomes after one feedback element as compared to no feedback as well as their evolution over the course of the feedback for the treatment groups only, i.e. from before the feedback to after the first feedback (positive or negative) until the second feedback (negative or positive). This can then also be compared to the evolution of these outcomes for the control group with no feedback.

Feedback is also followed up by questions about how people feel. Similarly, students are asked how they felt about the feedback in general after the exam. These measures can be used to analyze mechanisms of a potential effect. The post-exam questionnaire contains some further measures that could point towards potential mechanisms. These include the recall of feedback as a sanity check of the design as well as measures of how useful and difficult students thought the practice questions and the feedback were.

2.2. *Control variables*

A series of characteristics will be elicited from the respondents. They will be further explained in the order of appearance in the questionnaires.

1. **Sign up:**

- email address (only used for contacting and connecting responses)

2. **First questionnaire:**

- background characteristics:
 - age (integer values above 17, all ages above 25 will be categorized into *above 25* in the later analyses)
 - gender (options female, male, diverse)
 - semester (1 to 6 on integer scale, >6)
 - field of study (options are majors that offer the course)
 - proxy for ability: high school GPA and last high-school math grade (including option for non-German high school degree)
 - proxy for socio-economic status: each parent having attended college (yes/no/I don't know), each parents' current employment status (unemployed/employed/self-employed/retired/I don't know)

- proxy for culture: student and each parent born in Germany (yes/no/I don't know)
- personality traits, i.e. *How much do you agree with the following statements?* (not agree at all/rather not agree/neutral/rather agree/fully agree)
 - patience: *I am willing to give up something that is beneficial for me today in order to benefit more from that in the future., I prefer to have fun today and don't think about tomorrow.* (reversed), *I tend to postpone things until later, even if it would be better to do them immediately.* (reversed) (see [Resnjanskij et al. \(2021\)](#))
 - risk aversion: *In general, I am willing to take risks., I try to avoid risky situations.* (reversed) (similar to [Falk et al. \(2018\)](#))
 - big 5 (as in [Resnjanskij et al. \(2021\)](#))
 - * conscientiousness: *I am someone who tends to be lazy.* (reversed); *I am someone who does a thorough job.*
 - * neuroticism: *I am someone who is relaxed, handles stress well.* (reversed); *I am someone who gets nervous easily.*
 - * openness: *I am someone who has few artistic interests.* (reversed); *I am someone who has a vivid imagination/fantasy.*
 - * extraversion: *I am someone who is reserved.* (reversed); *I am someone who is outgoing, sociable.*
 - * agreeableness: *I am someone who is generally trusting.; I am someone who tends to find fault with others.* (reversed)
 - self efficacy (as in [Resnjanskij et al. \(2021\)](#)): *In difficult situations, I can trust in my abilities.; I am able to solve most problems on my own.; I can usually solve even challenging and complex tasks well.*
 - feedback aversion: *I generally like to get feedback on my performance.* (reversed), *Getting others' opinion on my performance helps me improve.* (reversed), *It makes me anxious to hear what others think about my performance.*

The score for all of these measures will be an equally weighted average of the single elements.

3. **Feedback:** *How do you feel about the feedback you just received?* (very bad/bad/neutral/good/very good)

4. **Post-exam questionnaire:**

- actual study plan (see above)
- difficulty of exam (very difficult/rather difficult/neutral/rather easy/very easy)
- difficulty of respective practice questions as compared to exam (much harder/a bit harder/similar/a bit easier/much easier)
- usefulness of respective practice questions for exam preparation (not useful at all/rather not useful/neutral/rather useful/very useful)
- usefulness of feedback (not useful at all/rather not useful/neutral/rather useful/very useful)
- feedback recall (customized depending on treatment group)

3. Hypotheses/Analysis

The structure of the treatment groups implies various possible treatment effects to be studied. The first and most important effect in this paper is the difference in outcomes between the groups receiving two feedback elements in different order. This comparison will show how the ordering of positive and negative feedback elements affects the perception and consequences of feedback. The main specification will look as follows. It excludes the control group and only looks at differences between the two feedback orders.

$$O_i = \beta_0 + \beta_1 POSNEG_i + \mathbf{X}_i + \varepsilon_i \quad (1)$$

O_i refers to the different outcomes as described above. $POSNEG_i$ refers to the treatment group with positive feedback first and negative feedback afterwards, the reference group is the negative first treatment. ε_i is an individual error term. I use robust standard errors. The matrix $\mathbf{X}_i$ refers to potential controls to be included. These consist of the background characteristics described above (age, gender, semester, field of study, high-school GPA, last math grade, high-school math grade, parents' education and occupation, migration status) as well as the personality traits that are also elicited prior to the first set of practice questions (patience, risk-aversion, big 5, self-efficacy, feedback aversion). Lastly, it includes dummies for the respective course-gender cell within which I randomize as well as for the topics students got feedback on. Further specifications will also include performance controls from the practice questions. Here, an interesting comparison refers to the differences in the immediate performance in the second set of practice questions as compared to the medium-term but high-stake outcome in the exam.

To understand the evolution of motivation and beliefs over the cycle of the feedback, the following specification will be used, including the control group as the baseline category:

$$\begin{aligned} O_{i,t} = & \beta_0 + \beta_1 initial_POSNEG_{i,t} + \beta_2 firstfb_POSNEG_{i,t} + \beta_3 secondfb_POSNEG_{i,t} \\ & + \beta_4 initial_NEGPOS_{i,t} + \beta_5 firstfb_NEGPOS_{i,t} + \beta_6 secondfb_NEGPOS_{i,t} \\ & + \beta_7 after_CONTROL_{i,t} + \mathbf{X}_i + \varepsilon_{i,t} \end{aligned} \quad (2)$$

Hereby, the initial beliefs and motivation of the control group will be standardized to 0 such that all other measures express deviations from the initial control mean. Standard errors will be clustered at the individual level, all else is equivalent to equation (1). This comparison with the control group receiving no feedback after each of the feedback rounds can shed light on the effects of no feedback v. one feedback (first round) as well as no feedback v. two feedbacks (second round). For performance, this cannot be done equivalently since there is no performance measure between feedback elements.

In addition to these average treatment effects, some heterogeneities are theoretically of interest. According to prior literature, it is very likely to see different results by gender. Women seem to react more heavily to negative feedback (see e.g. [Coffman et al. \(2021\)](#)) which in this setting would imply that the feedback order with negative feedback first could have an impact on how they perceive the later positive feedback. For men, this could even be reversed considering evidence of them being (more) overconfident (see e.g. [Niederle and Vesterlund \(2007\)](#)). Furthermore, there might be heterogeneities along the performance distribution. Low-performers might be especially motivated by (early)

positive feedback if it is something they rarely receive. Equivalently, this could hold for high-achievers and (early) negative feedback. Heterogeneities could also be observed by culture (see e.g. [Shipper et al. \(2007\)](#); [Sully De Luque and Sommer \(2000\)](#)). To proxy for culture, students can be divided into natives (parents and participant born in Germany), first-generation migrants (parents and participant born abroad), and second-generation migrants (parents born abroad, participant born in Germany). Similarly, the socioeconomic status of the participants could matter if e.g. children of parents of different socioeconomic status are more used to certain feedback types/orders. In this study, all participants are university students such that there might not be a lot of variation in students' socio-economic background. The variation here will hence use dummies for both parents having a university degree or not. Finally, based on prior literature, it is in great interest of this paper to look closer at dimensions such as the big 5, patience and risk-aversion, self-efficacy, self-evaluation as well as feedback aversion as mechanisms for a potential effect.

To address a few of the potential issues of this study, some further robustness checks will be performed. Firstly, the distribution of final exam grades of the participants will be compared to the publicly available statistics for the exam grades of all course participants. A priori, it is not clear how students will select themselves into the experiment. It might both be the most motivated students who participate as well as those most in trouble with studying for this exam in order to receive further practice. Furthermore, the monetary incentive might be most important for low-performers as a compensation for the struggles with studying for the exam, but also for high-performers as a reward. Secondly, the attention dedicated to the background questionnaire can be assessed via the reverse coding of some of the questions on personal characteristics as described above. Furthermore, a robustness check will exclude the 1% quickest individuals in answering the background questionnaire. Balancing checks will be performed on all pre-collected background characteristics as well as the pre-treatment outcomes. Attrition will be analyzed mainly according to treatment status, but pre-treatment outcome measures as well as background characteristics will also be considered. Finally, there could be some minor interest in the choice of online platform for individuals who win one of the vouchers.

^a kinne@ifm.de, ifm Institute Munich

The author gratefully acknowledges generous funding by the Joachim Herz Foundation.

References

- Azmat, G., Bagues, M., Cabrales, A., and Iriberri, N. (2019). What you don’t know... can’t hurt you? a natural field experiment on relative performance feedback in higher education. *Management Science*, 65(8):3714–3736.
- Bandiera, O., Larcinese, V., and Rasul, I. (2015). Blissful ignorance? a natural experiment on the effect of feedback on students’ performance. *Labour Economics*, 34:13–25.
- Coffman, K. B., Araya, P. U., and Zafar, B. (2021). A (dynamic) investigation of stereotypes, belief-updating, and behavior. Technical report, National Bureau of Economic Research.
- Exley, C. L. and Kessler, J. B. (2022). The Gender Gap in Self-Promotion. *The Quarterly Journal of Economics*. qjac003.
- Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., and Sunde, U. (2018). Global evidence on economic preferences. *The Quarterly Journal of Economics*, 133(4):1645–1692.
- Hanushek, E. A., Kinne, L., Lergetporer, P., and Woessmann, L. (2022). Patience, risk-taking, and human capital investment across countries. *Economic Journal*, *forthcoming*.
- Hermes, H., Huschens, M., Rothlauf, F., and Schunk, D. (2021). Motivating low-achievers - relative performance feedback in primary schools. *Journal of Economic Behavior & Organization*, 187:45–59.
- Niederle, M. and Vesterlund, L. (2007). Do women shy away from competition? do men compete too much? *The quarterly journal of economics*, 122(3):1067–1101.
- Resnjanskij, S., Ruhose, J., Wiederhold, S., and Woessmann, L. (2021). Can mentoring alleviate family disadvantage in adolescence? a field experiment to improve labor-market prospects.
- Shipper, F., Hoffman, R. C., and Rotondo, D. M. (2007). Does the 360 feedback process create actionable knowledge equally across cultures? *Academy of Management Learning & Education*, 6(1):33–50.
- Sully De Luque, M. F. and Sommer, S. M. (2000). The impact of culture on feedback-seeking behavior: An integrated model and propositions. *Academy of Management Review*, 25(4):829–849.

Ich habe Ihre Antworten zu den Übungsaufgaben korrigiert und bewertet und möchte Ihnen nun ein personalisiertes Feedback dazu geben. Wie Sie vielleicht gemerkt haben, bezogen sich die Aufgaben auf verschiedene Themenbereiche aus dem Kurs Statistik I.

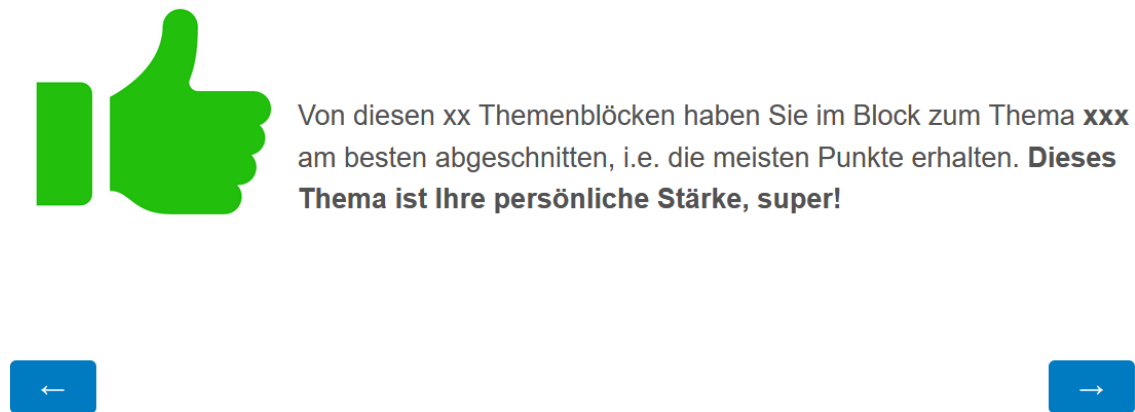


Figure 2: Screenshot of positive feedback provided to students

Ich möchte Ihnen nun ein weiteres Feedback zu Ihrer Leistung in den Übungsaufgaben vor wenigen Tagen geben. Dieses bezieht sich wiederum auf die eben angesprochenen Themenblöcke.

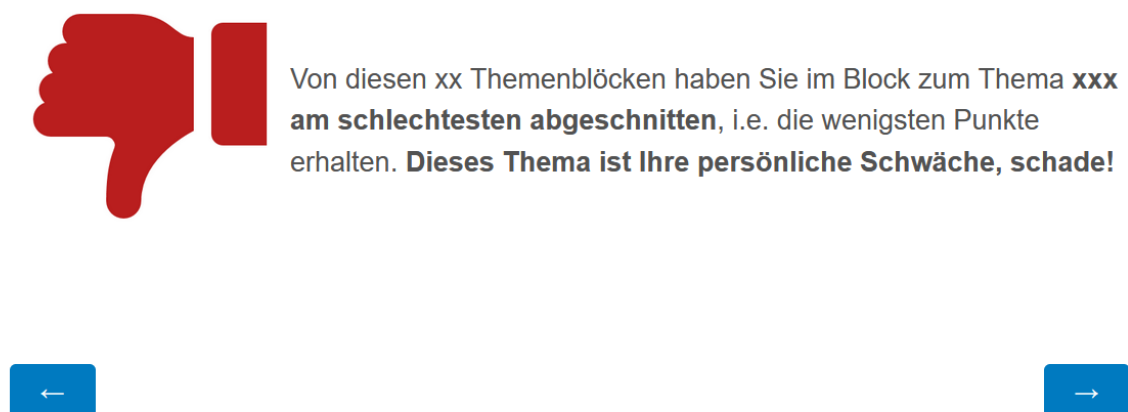


Figure 3: Screenshot of negative feedback provided to students