

Pre-Analysis Plan

Students and Their Thoughts

Prateek Bhan¹, Christina Felfe¹²³, Theodore
Koutmeridis⁴, Maike Schlosser¹, Judith Vornberger¹,
Jinling Wen⁴

October 2022

¹University of Würzburg, Würzburg, Germany

²Center of Economic Performance, London, United Kingdom

³CESifo, Munich, Germany

⁴University of Glasgow, Glasgow, Scotland

1 Introduction

This Pre-Analysis plan outlines the research design of a randomized controlled trial (RCT). With this study, we aim to examine the impact of mental health on academic performance and the role played by reflection and literacy intervention. In that attempt, primary data of students of an undergraduate course at the University of Würzburg will be collected during one academic semester.

During mid-20s, 75% of all mental health problems developing over life begin (Kessler et al., 2007). In particular, university students represent a vulnerable group with increasing mental health problems (Storrie et al., 2010). Poor mental health, especially depressive symptoms, have been linked to lower cognitive performance by a large body of literature (Cornaglia et al., 2015; Nafilyan et al., 2021; Eisenberg et al., 2009). Despite this correlation and the initial stage of disorder being a sensitive phase, there remains a massive care-seeking gap (Kessler et al., 2007). There exists growing evidence indicating that interventions during this sensitive period might help to prevent the duration and seriousness of problems (McGorry et al., 2011). Nevertheless, the potential impacts of soft touch preventative interventions remain largely unexplored.

We contribute to the existing literature by measuring both the impact of our intervention on mental health and the effect of mental health on academic performance. The results will test whether a reflection and/or a potential literacy intervention in the classroom can significantly affect mental health. We contribute to the literature by establishing a causal relationship for the effect of mental health on cognitive functioning within undergraduate students. Beyond that, the follow-up surveys will allow us to examine how stress develops over the course of an academic semester. This pattern could provide insights into when interventions are most effective.

Our study is moreover supposed to provide relevant findings for university life. Thus, high stress level, low life satisfaction, or mental health issues might already exist in second-year students. Rather than just measuring these determinants as a snapshot, our study observes students over the course of a semester. This allows us to understand how students' stress patterns develop and how the university can intervene. While we focus on German undergraduate students, all Higher Education Institutions (HEIs) can benefit from our findings. While the magnitude of certain factors may depend on the specific population being studied, our results could spawn further studies in different settings. These could in turn influence the design of taught degrees globally and thereby significantly impact wider society.

2 Research Strategy

2.1 Research Question

- Does reflection increase mental health and wellbeing?
- Does reflection increase students' effort?

- Does reflection increase students’ performance?

2.2 Sample

In total, the sample contains approximately 250-350 undergraduate students studying at the University of Würzburg in Germany and attending the undergraduate course taught by employees by the Chair of Labour Economics.⁵ These subjects are randomly assigned into two homogeneous and comparable groups: one treatment and one placebo group, with randomization at an individual level. To obtain approval of the competent parties⁶, the Chair of Labour Economics at the University of Würzburg was contacted in August 2022 and the University of Würzburg in early October 2022.⁷

2.3 Assignment to Treatment

For assignment to treatment (see Table 1), randomization is conducted in real-time at an individual level. For this, a QR code of the established web-based tool Qualtrics is presented to the participants, which then randomly allocates them to two homogeneous groups. To double blind the experiment, we add an active placebo group to the treatment group instead of a pure control group. Of course, the researchers do not know of these purely random treatment assignments either. Both of these groups receive a thought experiment using an identical text. Subsequently participants get a set of two questions depending upon their treatment assignment. Both are positively worded and asking for ways to better student wellbeing.

Treatment	Placebo
Thought experiment	Thought experiment
Questions concerning stress	Questions concerning architecture and aesthetics

Table 1: Assignment to Treatment

2.4 Intervention

Poor mental health is linked to lower cognitive performance (Cornaglia et al., 2015; Nafilyan et al., 2021; Eisenberg et al., 2009). Giving advice might increase motivation to study (Eskreis-Winkler, 2018). We request the students to

⁵The statistical power depends on the respective enrollment of students in the undergraduate course. Since we expect variation, we can not predict statistical power precisely.

⁶Institutional review boards (IRBs) approval was received from University of Glasgow of social sciences ethics committee in early October.

⁷Without revealing the exact outcome variables and treatment strategy to avoid any experimenter demand effects arising at an institutional level.

reflection by a thought experiment.⁸

The lab-in-the-field RCT is conducted in the lecture hall and is supervised by the lecturer and trained research assistants to avoid any possibility of interaction between students and to limit spill-over effects while completing the survey. During the completion only the assistants - independent from the course - check students' attendance to avoid possible experimenter demand effects. Randomization at an individual level is used to split the sample of participants into two homogenous groups. One treatment group and in order to avoid issues of compliance one placebo group. Both groups are asked to read an identical text including a thought experiment. The treatment group will answer two questions involving an identification of (i) the prevalent factors that may cause stress among university students and (ii) the ways to deal with these stressors to improve student wellbeing. The placebo group answers two questions of similar length related to architecture. These involving an identification of (i) important architectural factors for a university building and (ii) the ways to beautify the building.

The comparison of academic performance (exam grades) between the treatment group and the placebo group and evaluating students' development of mental health after the intervention enables us to examine the treatment effect of the thought experiment.

Overall, there are several reasons to assume a lower bound estimation of this effect. The students participating in the RCT are at least in their third semester and have thus likely already developed coping strategies to deal with stress during their first year of university. Unlike first year students, they have already taken several exams (5-6 exams are taken per semester) and are thus more experienced when the intervention takes place. The intervention could therefore be less potent for these experienced students compared to those in their first semester. In addition, post treatment communication after the lecture between students and friends makes treatment spill-over possible. While the environment can be controlled during data elicitation, participants can communicate after class. This will likely induce some individuals in the placebo group to reflect on the intervention received by the treatment group. Due to treatment spill-over and experience effect, any significant effects we may discover should be interpreted as lower bound estimates. This also highlights the cost effectiveness of the intervention if it can have small positive spill-over effects.

3 Outcome of Interest

The outcome of interest consists of two main categories: psychological factors and academic performance. Psychological outcome variables are measured using various established scales. Student performance is evaluated by attendance and

⁸Treatment group is asked what caused them stress in the past semester and its remedies. Placebo group is asked what architectural aspects of the university building appealed to them and how can those be beautified.

grades/scores over time, while performance is measured by two different test approaches. The individual instruments are listed and briefly described below:

1. Psychological factors

- **Depressive disorder** - This variable is measured via the eight-item Patient Health Questionnaire (PHQ-8). Each question includes four response options from 'Not at all' to 'Nearly every day' (Kroenke et al., 2009).
- **Stress** - is gauged using the perceived stress scale that includes ten questions with five response options each (from 'never' to 'very often') (Cohen et al., 1994).
- **Life satisfaction** - We include the question to rate life satisfaction on a scale from 0 ('completely dissatisfied') to 10 ('completely satisfied') used in Socio-Economic Panel (Kantar Public, 2021).
- **Risky behaviour** - For this we refer to the Busch et al. (2014) questions on the frequency of consuming stimulants (alcohol, cigarettes) during the last 30 days.

2. Performance

- **Attendance** is monitored during first lecture and during lectures with multiple-choice tests (in lectures 4, 6, 8, 10) by asking for the number of the participants.
- **Performance** - Firstly, this is assessed using the individual scores for all multiple choice test and the total score for all multiple choice tests. Secondly, the outcomes of the final exam and the re-sit exam are used. For the exams we can distinguish between drop-out (appearing, but cross-out; do not appear) and the evaluation of the exam (grade and individual scores).

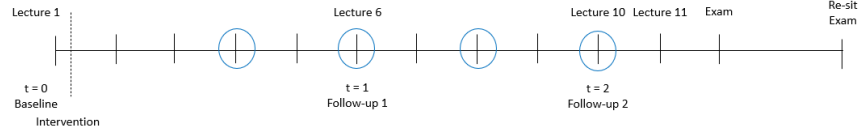
4 Data Collection

Data collection is conducted using Qualtrics over the course of an academic semester (from October 2022 to April 2023), as illustrated in Figure 1. The vertical black lines illustrate the semester structure including eleven lectures scheduled on a weekly⁹ basis, one final exam at the end of the semester and one re-sit exam in April 2023.

The project starts by administering the baseline survey in the first lecture (t=0 on October, 24 2022) to monitor the development over the semester and due to the expected high attendance in the first lecture. Here, confounding factors (such as gender, age, number of semesters previously studied) as well as initial mental health (e.g. depressive symptoms), general well-being (e.g. life satisfaction) and mindful behaviour (e.g. frequency of sport) of the students are

⁹With exceptions due to holidays.

Figure 1: Timeline



Note: The figure presents the timeline over the course of an academic semester from October 2022 to April 2023. It includes lectures taught (vertical black lines), intervention (dashed vertical black line), dates of the multiple choice tests (blue circles) and the timing of the survey waves (displayed below the timeline).

gathered. The participant group consists of the respondents of this baseline survey. The intervention takes place during Lecture 1 but after collecting the basic data (dashed vertical black line). Subsequently, both groups are asked about the percentage of answers that relied on Google. This enables to differentiate the role of reflection from additional or new information to identify a potential literacy intervention. A non-pecuniary incentive scheme (gift vouchers of a local market chain) will be put in place to reimburse students for the time that they will spend partaking in surveys and the intervention.¹⁰

As indicated below the timeline, two follow-up surveys will be collected. The first follow-up survey (t=1 on November, 28 2022) will take place in Lecture 6 and the second one (t=2 on January, 30 2023) in Lecture 10 to keep the intervals between the surveys similar.¹¹ While the baseline survey should take 20-25 minutes to fill out, the two subsequent surveys are shorter and will take less than 10 minutes. The first follow-up survey collect data on mindfulness and mental health, while the second one additionally includes questions on perceived stress, risky behaviour and life satisfaction. Participation in the two follow-ups is incentivized by a voucher lottery.

During the semester, four multiple choice tests (Lecture 4, 6, 8, 10, see blue circles in Figure 1) will be administered with questions related to the course topics covered up until that point in time. These tests enable to track attendance as well as performance during the semester. Attending the tests is incentives by offering a grade point jump after attending and also reaching 50% of the total score of all tests.

In the beginning of each questionnaire, during the multiple choice tests, and in the final and re-sit exams, students are asked for their matriculation numbers to link the data across the semester. Overall, the incentive scheme will act as a safety net to reduce attrition within and between groups. After collecting data,

¹⁰In baseline survey all participants receive a voucher irrespective of their assignment to treatment or control group.

¹¹We select these two occasions of follow-up surveys since the first one is in the middle of the semester and the second one takes place towards the end and right before the exam phase.

we use STATA for data analysis.¹²

5 Empirical Analysis

Empirical Strategy

We estimate the effect of the reflection intervention on college students, using the following equation by Ordinary Least Squares (OLS):

$$Y_i = \alpha_1 + \beta_1 Treatment_i + \varphi X'_{i0} + \varepsilon_i \quad (1)$$

where Y_i is the outcome of interest for student i such as mental illness and academic performance. $Treatment_i$ is a binary indicator equal to one if student i is assigned to the treatment group. X'_{i0} is a vector of baseline characteristics for student i such as gender, age, and among others.¹³ ε_i is the error term. Robust standard errors are calculated to allow for heteroskedasticity. β_1 captures the effects of the reflection intervention on students. We assess the intervention effect immediately, in the middle of the term and in the end of the term after the intervention depending upon the outcome variable and the instrument.

To investigate our research questions, we test the following hypothesis:

Ho : The reflection intervention has no effect on the outcome variable, $\beta_1 = 0$

Following McKenzie (2012), we also estimate the treatment effects using an ANCOVA specification of the form:

$$Y_{it} = \alpha_2 + \beta_2 Treatment_i + Y_{i0} + \varphi X'_{i0} + \varepsilon_i \quad (2)$$

where Y_{it} is the outcome of interest for student i in time t .¹⁴ $Treatment_i$ is a binary indicator equal to one if student i is assigned to the treatment group. Y_{i0} is the baseline value of the outcome. X'_{i0} is a vector of baseline characteristics for student i such as gender, age, and among others. ε_i is the error term. Robust standard errors are calculated to allow for heteroskedasticity.

To answer our research questions, we have the following hypothesis to test:

Ho : The reflection intervention has no effect on the outcome variable, $\beta_2 = 0$

Heterogeneous Effects

We collect information on various demographic characteristics or basic data such as age, gender, number of semesters previously studied, and current grade point average in the baseline survey.¹⁵ Since we expect possible heterogeneous

¹²After matching data, we generate a unique ID and drop the matriculation number to have a totally anonymous dataset.

¹³We follow the common practice of adjusting and presenting the results with the inclusion of baseline characteristics. We explore the sensitivity of the results to the exclusion of covariates.

¹⁴ANCOVA is only used for outcomes measured both before and after the intervention.

¹⁵Additionally, we may obtain further heterogeneous treatment effects through the range of information available once the data collection is completed.

treatment effects across those confounding characteristics, examining differential treatment effects will be of interest.

Robustness Checks

- We will estimate both Equation (1) and (2) of Section ?? without including the vector of controls X_{i0} .
- Associations at baseline will be computed to control for these when estimating treatment effects via Ordinary Least Squares (OLS) and ANCOVA.
- The formation of clusters is already prevented by a single treatment at the individual level.

6 Organisation and Contribution

The project aims at assessing the impact of mental health on academic performance. The investigators are based at the University of Würzburg and the University of Glasgow. Research costs of the project are covered by the Chair of Labour Economics of the University of Würzburg. The results will extend existing research on reflection, academic performance and mental health. While we focus on German undergraduate students, all HEIs can benefit from our findings. Our results could spawn further studies in different settings and could in turn influence the design of taught degrees globally and thereby significantly impact wider society.

References

- [1] Busch, S. H., Golberstein, E., & Meara, E. (2014). *The fda and abcs unintended consequences of antidepressant warnings on human capital*. Journal of Human Resources, 49(3), 540-571.
- [2] Cohen, S., Kamarck, T., & Mermelstein, R. (1994). *Perceived stress scale. Measuring stress: A guide for health and social scientists*, 10(2), 1-2.
- [3] Cornaglia, F., Crivellaro, E., & McNally, S. (2015). *Mental health and education decisions*. Labour Economics, 33, 1-12.
- [4] Eisenberg, D., Golberstein, E., & Hunt, J. B. (2009). *Mental health and academic success in college*. The BE Journal of Economic Analysis & Policy, 9(1).
- [5] Eskreis-Winkler, L., Fishbach, A., & Duckworth, A. L. (2018). *Dear Abby: Should I give advice or receive it?*. Psychological science, 29(11), 1797-1806.
- [6] Kantar Public (2021). *SOEP-Core – 2020: Individual (A-L3, M1-M2 + N-Q)*. SOEP Survey Papers 1069: Series A. Berlin: DIW/SOEP.
- [7] Kessler, R. C., Amminger, G. P., Aguilar-Gaxiola, S., Alonso, J., Lee, S., & Ustun, T. B. (2007). *Age of onset of mental disorders: a review of recent literature*. Current opinion in psychiatry, 20(4), 359.
- [8] Kroenke, K., Strine, T. W., Spitzer, R. L., Williams, J. B., Berry, J. T., & Mokdad, A. H. (2009). *The PHQ-8 as a measure of current depression in the general population*. Journal of affective disorders, 114(1-3), 163-173.
- [9] McGorry, P. D., Purcell, R., Goldstone, S., & Amminger, G. P. (2011). *Age of onset and timing of treatment for mental and substance use disorders: implications for preventive intervention strategies and models of care*. Current opinion in psychiatry, 24(4), 301-306.
- [10] McKenzie, D. (2012). *Beyond Baseline and Follow-up: The case for More T in Experiments*. Journal of development Economics, 99(2), 210-221.
- [11] Nafilyan, V., Avendano, M., & de Coulon, A. (2021). *The causal impact of depression on cognitive functioning: evidence from Europe*. Available at SSRN 3771732.
- [12] Storrie, K., Ahern, K., & Tuckett, A. (2010). *A systematic review: students with mental health problems—a growing problem*. International journal of nursing practice, 16(1), 1-6.