MIDS Slacker Project: When Should You Slack your MIDS Instructor?

Amanda D'Alessio, Darya Likhareva, Meir Marmor, Noor-Ul-Ain Al

UC Berkeley School of Information

W 241, Fall 2023: Experiments and Causality

Abstract

Background: The UC Berkeley Masters of Information and Data Science program (MIDS) relies heavily on electronic communication due to the program's online format. The Slack instant messaging app is the primary communication method in the program. Therefore, optimal use of this platform is imperative for successful participation in the MIDS program. This study aimed to identify the optimal time to send a Slack message to MIDS course instructors.
Methods: MIDS instructors from all courses were randomly assigned to receive identical Slack messages during work hours (Mon-Fri, 9 a.m. to 5 p.m. PST) or outside work hours (Mon-Fri, 5 p.m. to 9 a.m. PST). Response time was the primary outcome measure, the time between sending the message and receiving any response. The experiment was repeated with four different messages four times to assess the influence of message type on response time.
Results: Using a response window of 48 hours, our average response rate was 55% (52-60%). Regression analysis with clustered standard errors showed that sending a message within working hours resulted in a reduction in 4 hours of response time ($p<0.05$). Individual message analysis showed a statistically significant decrease in response time by 5 hours, only for message 1. An F-test confirmed the superiority of the message 1 model. Whether the instructor teaches a foundational course did not significantly influence response time in combination with any of the messages. However, our experiment was severely underpowered, and a higher sample size may have resulted in significant results for more of the messages.
Conclusions: Our findings suggest that students should aim to send their message during working hours, mainly if it involves a simple task for the instructor. To answer our study question more definitively, the experiment needs to be repeated with a larger sample size and the collection of more demographic data.

Background

The UC Berkeley Master of Information and Data Science (MIDS) program is an online master's program that relies heavily on electronic communication. Slack is a cloud-based team communication platform developed by Slack Technologies and currently owned by Salesforce. The Slack app is very commonly used by students, alums, and faculty of the UC Berkeley School of Information, making it the primary method of communication in the MIDS program. Due to the program’s online nature, optimal use of the Slack app may improve the MIDS students'

academic experience. Based on the personal experience of the investigators, MIDS instructors' response time to Slack messages during MIDS courses is highly variable and occasionally frustrating to students. We hypothesized that the sending time of these messages significantly influences their response rate.

While there is limited research on instant messaging response rates in academic settings, existing studies focusing on messaging response etiquette offer valuable insights. For instance, the study by Aranda and Biag in 2018 [1] reveals that recipients of messages often feel compelled to respond quickly, ideally within a timeframe ranging from 20 minutes to the end of the day, to adhere to social etiquette and avoid offending the sender. Additionally, we can draw parallels from studies on SMS marketing campaigns to understand response dynamics further. Although the context of these studies differs from student-instructor interactions, they provide a broader perspective on typical response behaviors. As highlighted in EZ Texting's 2019 report on Mobile Usage [2], a significant proportion of individuals read texts within five minutes of receipt, and many respond within three minutes. This indicates a response time variability spanning 3 to 20 minutes, potentially mirroring expected behaviors in academic settings.

Our experiment specifically aimed to identify the optimal times for MIDS students to send messages to their instructors. By examining these aspects, our research seeks to develop actionable strategies for MIDS students to enhance communication with their instructors, thereby contributing to the broader understanding of digital communication etiquette in online educational environments.

Research Question

The primary question we aim to answer through our research is whether MIDS instructors respond faster to Slack messages during Pacific Time working hours (9 AM to 5 PM) than outside these hours. This question can inform understanding of the digital communication dynamics in an academic setting, particularly the increased reliance on virtual platforms for education and instructor-student interactions.

Null Hypothesis (H0): There is no statistically significant difference in the response time of MIDS instructors when Slack messages are sent during work hours compared to after work hours. Instructors' responsiveness is not influenced by a typical workday schedule, suggesting that their engagement with digital communication platforms transcends typical office hours.

Alternate Hypothesis (H1): There is a statistically significant difference in the response time of MIDS instructors when Slack messages are sent during work hours compared to after work hours. It could indicate that instructors are more responsive during traditional working hours, potentially due to a higher engagement with school-related work during a traditional workday.

Understanding these dynamics has several practical implications. Firstly, it could guide students on the optimal times to expect a quicker interaction with their instructors. Secondly, it may provide insights into instructors' work habits and time management strategies in a remote or digital school environment.

Our study's findings could serve as a stepping stone for more extensive research into digital communication etiquette and strategies in the evolving landscape of online education, particularly as the online education sector grows. By defining the factors influencing response times, universities can tailor their student-instructor communication guidelines to align with the observed patterns, ultimately fostering a more efficient and responsive educational system.

Experimental Methods

Experiment Overview: In our experiment, we implemented a randomized control trial (RCT) to understand the response patterns of MIDS instructors to Slack messages. The setup was straightforward: subjects were randomly assigned to either the treatment or control group.

Post-experiment, we aimed to collect demographic data through a Qualtrics survey, seeking to supplement our analysis with a more robust background of instructors beyond their online response times. As described later in the limitations section, we struggled to get responses, so we ultimately relied on demographic information provided by the university (for example, is the course foundational?).

Our key metric for analysis was the average treatment effect, calculated as the difference in response times between the treatment and control groups (as seen above in the message example). This measure was essential to determine if the timing of message delivery significantly influenced instructors' response times.

Recruitment Process: For our experiment, we specifically targeted MIDS instructors actively teaching in the Fall 2023 semester. Our recruitment process began by obtaining an up-to-date list of faculty members from the MIDS enrollment department for that semester. This step was crucial to ensure that we included a representative sample of current instructors in our study. While we considered recruiting graduate teaching assistants for our study, we struggled to find a comprehensive list of active TAs. Our only method of getting an active and comprehensive TA list was by directly reaching out to professors, which would break the experiment.

Once we had compiled the list of active instructors, the next phase involved reaching out to these instructors. We chose Slack as our communication medium, as it is the standard platform MIDS

professors use for student interactions. By using Slack, we aimed to replicate how instructors and students typically receive and respond to messages.

The Slack platform was a convenient tool for sending messages as part of the treatment and control conditions and a way to track real-time responses. The instructors' familiarity with Slack was expected to reduce any potential learning curve or response bias from using an unfamiliar communication platform.

Primary outcome: Response time, defined as the time an instructor takes to respond to the original message.

Pilot data: Our team decided to run a pilot experiment to assess the efficiency of the general research design.

To simulate the experiment as closely as possible without explicitly using our sample, we ran an observational study collecting message send and response times across different Slack channels. We collected response times to 100 Slack messages in 4 different Slack channels:

Amanda	datasci-241-2023-fall
Darya	#mids-class-recs
Noor	#no-stupid-questions
Meir	#datasci-200-python

We aimed to group our messages as closely to our broad message categories as possible. With that in mind, the following message types and their subsequent counts were identified:

Question Types		Count
1	Task request	69
2	Life advice	6
3	Outside of the job description	9
4	Feedback request	16

As with our final experiment, we analyzed the responses based on 9AM - 5PM response times in the Pacific Standard Time zone. We identified 40 messages within work hours and 60 outside of working hours. Our analysis indicated that the response in working hours was shorter by an average of 1:35 hours:

WorkHours	AVERAGE of TimeToResponse
0	4:11:17

1	2:36:05
Grand Total	**3:33:12**
ATE =	1:35:12

We also used this pilot data and findings to conduct a power analysis (more details in the power analysis section).

Project Timeline:

The Fall 2023 term included two breaks – a Fall break and a Thanksgiving break. Recognizing that these periods might influence the typical communication patterns of instructors, we adjusted our messaging schedule to align with the academic calendar. Our goal was to optimize the response rate by avoiding periods when instructors were less likely to be engaged in their regular teaching and communication routines.

Messaging started on November 13th, with a plan to avoid sending messages during the breaks. This approach was intended to minimize potential disruptions in response behavior that could skew the study's findings. The final message was recorded on December 6th, ensuring we captured a representative sample of response times across the active weeks of the semester.

Instructional Week	Monday	Tuesday	Wednesday	Thursday	Friday
Class Registration	**Nov 13:** Send Message 1		**Nov 15:** Send Message 2		
Thanksgiving Break	**Nov 20:** Record Message 1		**Nov 22:** Record Message 2		
Week 12	**Nov 27:** Send Message 3		**Nov 29:** Send Message 4		
Week 13	**Dec 4:** Record Message 3		**Dec 6:** Record Message 4		

To ensure a comprehensive analysis, we tracked responses to each message for up to one week from its original send date. For instance, messages dispatched on November 13th were monitored until November 20th for any incoming responses. This one-week window was chosen to strike a balance between allowing ample time for instructors to respond and maintaining the relevance of the response to the original message.

Comparison of Potential Outcomes (ROXO diagram):

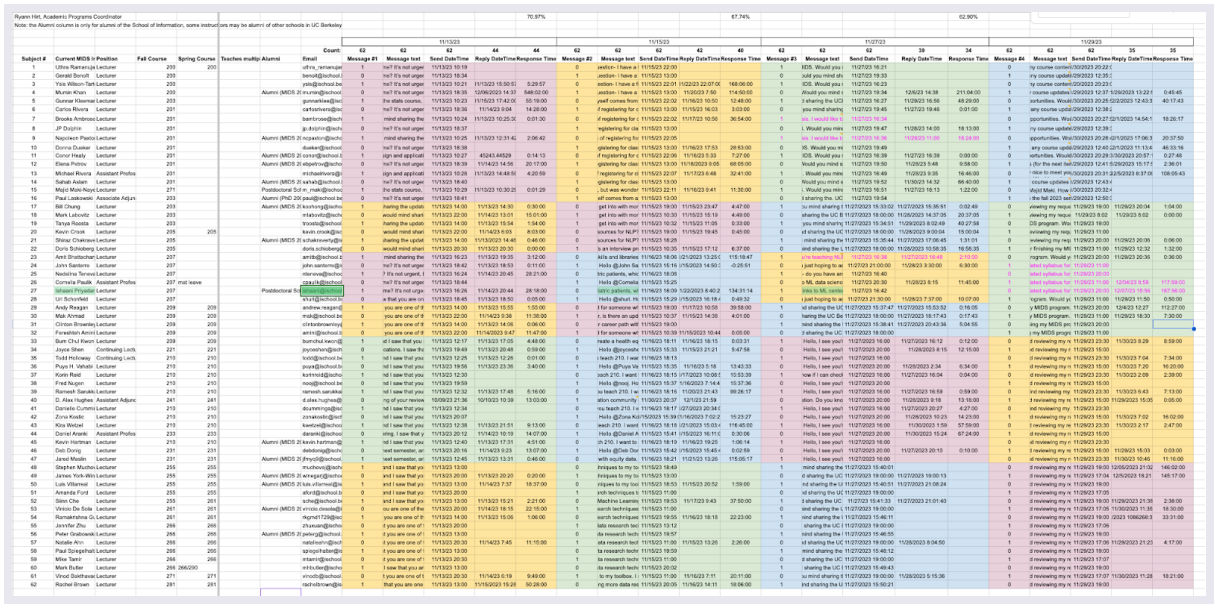
In our experiment, we have our group of four students sending out messages. The R in our diagram below, or the randomized experiment, only had two parts. The control section was a collection of response times of professors in prior conversations in various Slack group chats.

The O, observation, was the response times. The treatment was when our team sent out different messages both during work hours and outside of them and measured response rates, X being the applied treatment of working hours. The observation of this study was the response time, and we also included what kind of message was sent to get that response time.

N_{Amanda}	$R_{control}$ $R_{treatment}$	$O_{response\ time}$ $O_{response\ time}$ $O_{response\ time}$	$X_{working\ hours}$ $X_{non\ working\ hours}$	$O_{time,\ message\ type}$ $O_{time,\ message\ type}$ $O_{time,\ message\ type}$ $O_{time,\ message\ type}$
N_{Meir}	$R_{control}$ $R_{treatment}$	$O_{response\ time}$ $O_{response\ time}$ $O_{response\ time}$	$X_{working\ hours}$ $X_{non\ working\ hours}$	$O_{time,\ message\ type}$ $O_{time,\ message\ type}$ $O_{time,\ message\ type}$ $O_{time,\ message\ type}$
N_{Darya}	$R_{control}$ $R_{treatment}$	$O_{response\ time}$ $O_{response\ time}$ $O_{response\ time}$	$X_{working\ hours}$ $X_{non\ working\ hours}$	$O_{time,\ message\ type}$ $O_{time,\ message\ type}$ $O_{time,\ message\ type}$ $O_{time,\ message\ type}$
N_{Noor}	$R_{control}$ $R_{treatment}$	$O_{response\ time}$ $O_{response\ time}$ $O_{response\ time}$	$X_{working\ hours}$ $X_{non\ working\ hours}$	$O_{time,\ message\ type}$ $O_{time,\ message\ type}$ $O_{time,\ message\ type}$ $O_{time,\ message\ type}$

Randomization

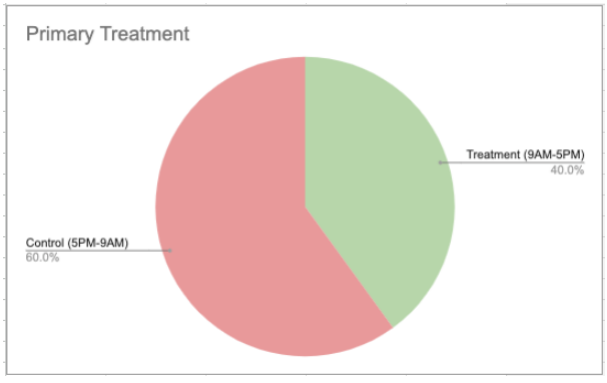
A twofold randomization scheme was employed. The instructors were randomized to receive either treatment (message sent during working hours) or control (message received after working hours by listing the 62 participating instructors in random order and then assigning treatment by alternating treatment and control assignment sequentially (1-0-1-0-1-...). A second level of randomization was achieved by sequentially alternating the message type and the messenger's identity so that the same message by the same messenger in the same order would be shared by only 25% of the participating instructors (see below).



As seen in the visual representation of the randomization scheme above. On the left, the participating instructors are listed in random order. The four colors represent the four different messengers. Treatment and control are alternated in each row.

Treatment

This study's primary treatment was sending a Slack message during working hours, defined as the period between 9 AM and 5 PM Pacific Standard Time. A message sent outside of these hours (anything outside of 9 AM to 5 PM) was designated control. We chose to delineate our study this way based on general societal conventions of a typical workday, and this was confirmed through consensus among the research team. Pilot data confirmed that this cut-off was a potential differentiator of response time.

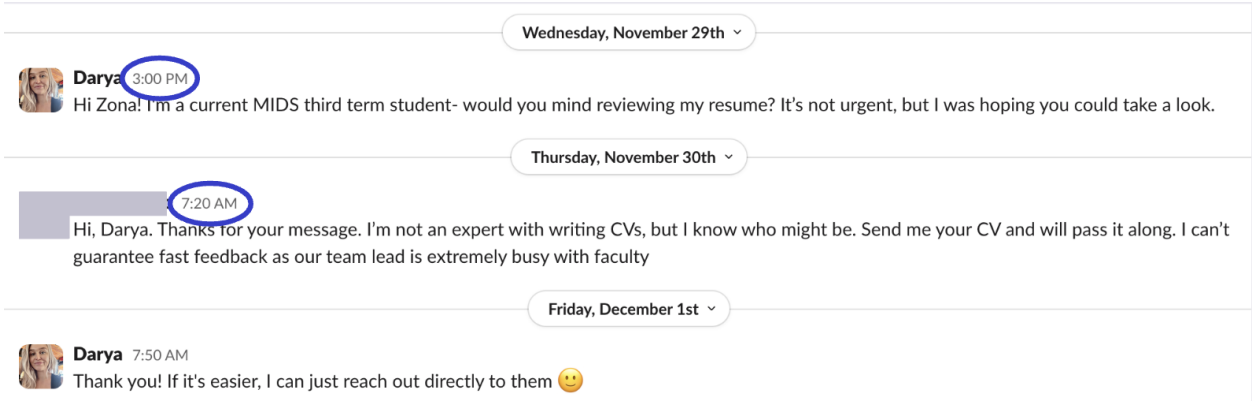


The primary treatment was administered four times, each time with a different message, and each time alternating instructors between treatment and control. The four messages represented four different anticipated response times and were sent to the participating instructors over four weeks

(see below). A shorter response time was anticipated for messages 1 & 3 and a longer response time for messages 2 & 4. It is important to note that we were looking for a written response to our message - so for the purposes of this study, we did not consider an emoji reaction a response.

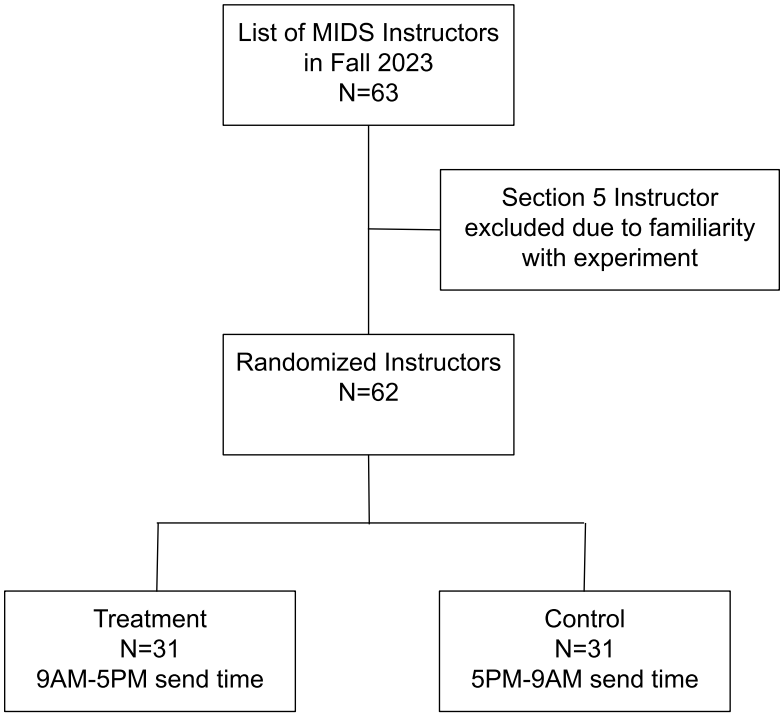
Message #	Message Type	Message Theme
1	Task request	I am interested in taking the course XXX, and I saw that you are one of the instructors for the course. Would you mind sharing the syllabus?
2	Life advice	Do you think this course would be good for a product data scientist?
3	Outside of their job description	Would you mind sharing the UCB Interviewing Resources doc?
4	Feedback request	Would you mind reviewing my resume? It's not urgent, but I was hoping you could take a look.

While we did adapt the message types to reflect each individual's "realistic" request, we aimed to keep within the themes of the message types. Below is an example of a message sent in the experimental phase:



Once we received a response to a message, we responded back to the instructor, either following through with the conversation or thanking them for their response. This was a crucial part of our experimental design, aimed at keeping engagement and realism in our interactions. We wanted to make sure our messages and responses mirrored a typical Slack conversation. By doing so, we aimed to prevent any potential bias that might arise if the instructors suspected they were part of a study.

CONSORT Diagram



The CONSORT diagram represents the general design of the experiment. The same design was repeated four times using four different messages and alternating the instructors between treatment and control.

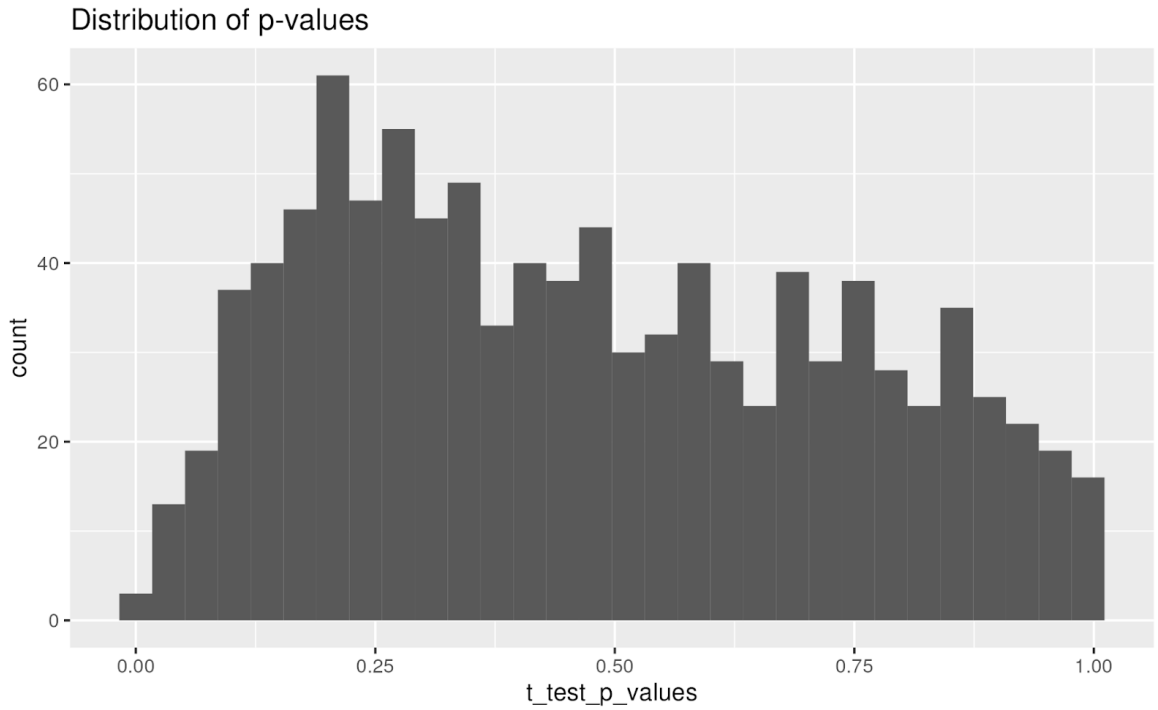
Power Calculation

In preparation for our study, we conducted a power analysis to determine the sample size needed to detect a statistically significant effect. We decided to use our pilot data to perform the power calculation.

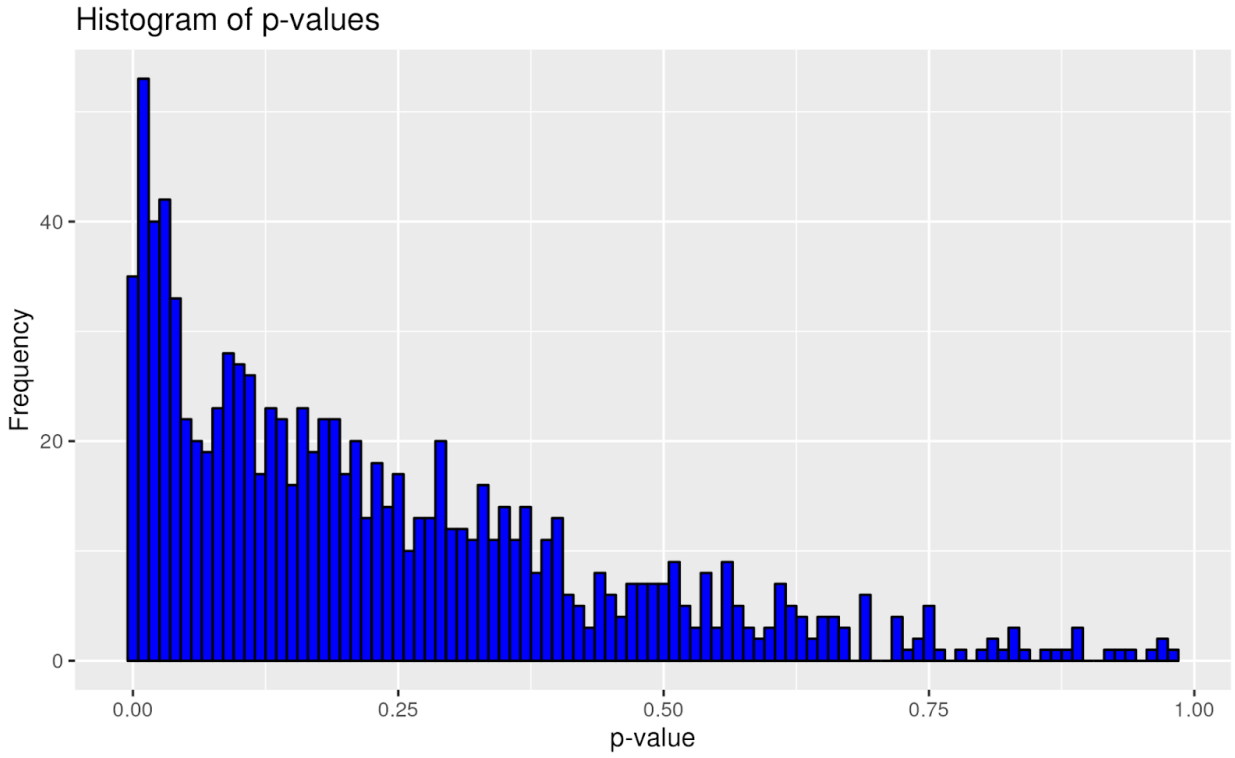
To perform a power analysis, we collected pilot data from the main MIDS slack channel of 100 pairs of messages and replies (50 during work hours and 50 after work hours) and created a 2-column data table of time_sent and response_time.

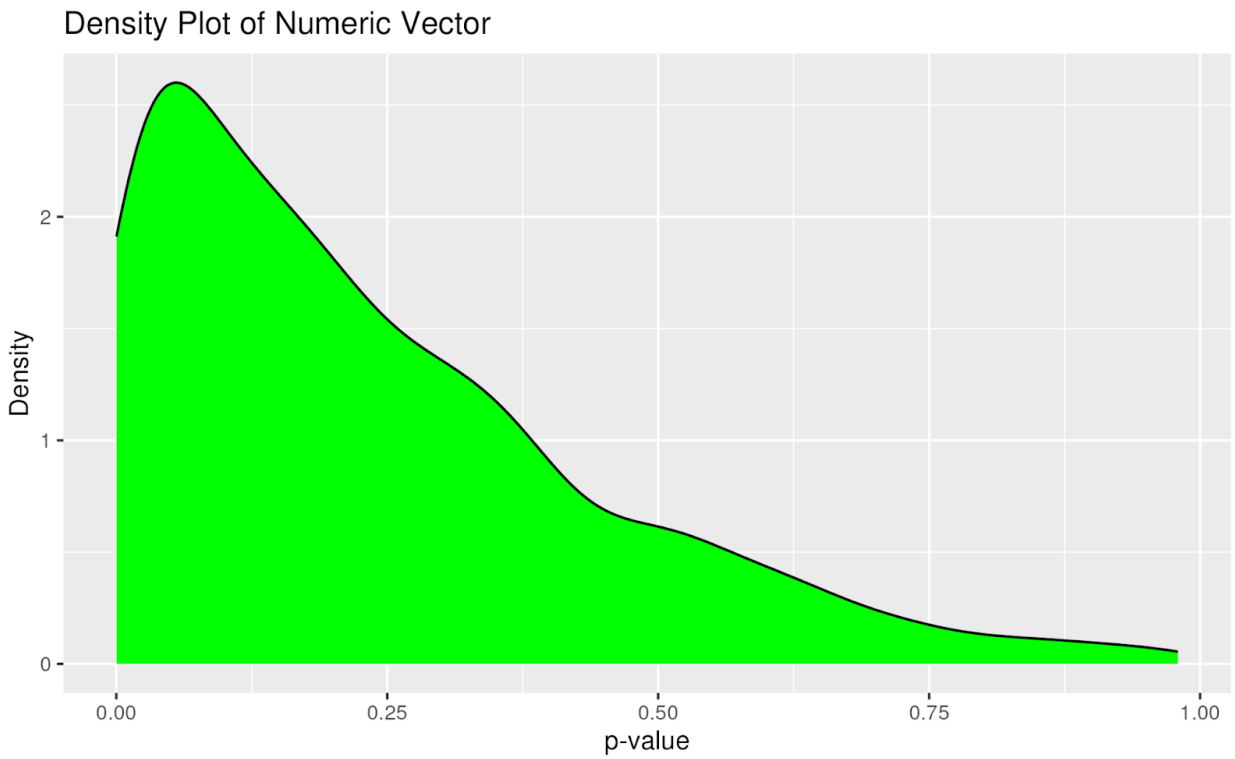
We sampled ten people from our existing data and conducted the appropriate t-test one thousand times. We replicated a few different scenarios with different defined working and non-working hours, with the scenario we tested in our experiment highlighted below.

Scenario 2 - Work Hours 9-5



With only ten subjects, it is not well-powered.





For a sample size of 40, with 20 subjects in the treatment group and 20 subjects in the control group, the experimental power is 0.184. Power generally increases as sample size increases. However, our field experiment was limited by the number of instructors participating in the fall semester. Therefore, the final sample size of 62 instructors severely underpowered our experiment to detect a difference in response time. If allowed more time to continue the experiment, we planned to expand the experiment to TAs and graders and increase our sample size.

Results

We used an Ordinary Least Squares (OLS) regression analysis to assess the relationship between message sending and response times. A response window of 48 hours was chosen to maximize the effect of message timing (e.g., responses that happened five days after the message was sent were less likely to be affected by the message sending time compared to responses that happened a day after the message was sent).

Analyzing the Impact of the Messages in Aggregate

Our primary analysis reveals a 55% response rate within a 48-hour time frame post-message dissemination (Figure). To account for a possible clustering effect of groups of instructors responding to messages faster than other groups (e.g., instructors from the same MIDS course could have a policy of responding to Slack messages within 48 hours), we made every instructor

represent a unique cluster. We calculated clustering standard errors by the instructor variable, creating 62 different clusters in our analysis.

Message 1 RR:	Message 2 RR:	Message 3 RR:	Message 4 RR:
60%	53%	56%	52%

Figure: Response Rate (RR) of the four different messages.

Regression analysis with clustered standard errors resulted in message-sending time statistically influencing response time. Specifically, sending a message within working hours resulted in a reduction in 4 hours of response time ($p<0.05$). Being an instructor in a foundational course (defined as those courses you take at the beginning of the program, like 201, 203, etc.) did not significantly affect response time and rejected the hypothesis that instructors working with students in their first or second terms, will be more sensitive to student messages and respond sooner.

As discussed later in the limitations section, adding more independent demographic variables would have helped us build a more robust analysis. However, as mentioned earlier, we sent out a follow-up survey to collect demographic data after the experiment closed but struggled to collect demographic data.

```
Combined Message 1,2,3 & 4 w/ Clustered SE
================================================
                        Dependent variable:
                    ----------------------------
                           response_time
------------------------------------------------
message                        -4.252**
                               (1.760)

foundational_course             0.896
                               (1.804)

Constant                       10.985***
                               (1.495)

------------------------------------------------
Observations                     137
R2                              0.043
Adjusted R2                     0.029
Residual Std. Error       10.300 (df = 134)
F Statistic             3.031* (df = 2; 134)
================================================
Note:                *p<0.1; **p<0.05; ***p<0.01
Analysis of Variance Table

Response: response_time
                     Df  Sum Sq Mean Sq F value  Pr(>F)
message               1   615.8  615.83  5.8043 0.01735 *
foundational_course   1    27.3   27.31  0.2574 0.61272
Residuals           134 14217.1  106.10
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Analyzing the Impact of Each Message Individually

Our analysis of each of the messages individually suggests that the timing of Slack messages (during vs. outside work hours) can influence the response times of MIDS instructors, with some evidence of faster responses during work hours for some specific messages.

As a reminder, here are the broad message categories:

Message #	Message Type
1	Task request
2	Life advice
3	Outside Job Description
4	Feedback request

Our analysis shows varied impacts of messages on response times. For Message 1, we observed a statistically significant decrease in response time by 5 hours. This is particularly insightful considering the context of the first message, which was sent during the week of registration. Instructors were probably hyper-aware of messages as a whole, especially when it came to questions about the syllabus. However, this is not the case for the rest of the messages, which have no statistically significant change in response time (messages 2, 3, and 4). Interestingly, being a foundational course instructor does not consistently affect response times across all models, as indicated by the non-significant coefficients for foundational_course.

	Dependent variable:			
	message_1_response_time	message_2_response_time	message_3_response_time	message_4_response_time
	(1)	(2)	(3)	(4)
message_1	-5.793***			
	(2.072)			
message_2		-8.283*		
		(4.234)		
message_3			-4.962*	
			(2.766)	
message_4				1.977
				(4.667)
foundational_course	-0.626	-2.512	4.428	2.060
	(2.103)	(4.266)	(2.766)	(4.667)
Constant	10.215***	17.451***	7.753***	8.871**
	(1.688)	(3.699)	(2.372)	(4.042)
Observations	37	33	35	32
R2	0.193	0.118	0.149	0.013
Adjusted R2	0.146	0.059	0.096	-0.055
Residual Std. Error	6.251 (df = 34)	12.074 (df = 30)	8.176 (df = 32)	13.201 (df = 29)
F Statistic	4.074** (df = 2; 34)	2.009 (df = 2; 30)	2.808* (df = 2; 32)	0.187 (df = 2; 29)
Note:				*p<0.1; **p<0.05; ***p<0.01

Figure: Regression analysis models for individual message control for foundational courses.

The F-tests conducted for each of the four regression models provide additional insights into the overall significance of the models. In message 1, the F-test result is statistically significant. The F-tests show lower significance levels for messages 2, 3, and 4, suggesting that message timing was more important for message one than for other messages. Whether the instructor teaches a foundational course was not significant in combination with any of the messages. It is, however, noteworthy that our experiment was severely underpowered, and a higher sample size may have resulted in significant results for more of the messages.

F-test Results:

```
Model 1 F-test:
Analysis of Variance Table

Response: message_1_response_time
                   Df  Sum Sq Mean Sq F value   Pr(>F)
message_1           1  314.92 314.920  8.0596 0.007585 **
foundational_course  1    3.46   3.456  0.0884 0.767978
Residuals          34 1328.52  39.074
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Model 2 F-test:
Analysis of Variance Table

Response: message_2_response_time
                   Df Sum Sq Mean Sq F value  Pr(>F)
message_2           1  535.1  535.12  3.6708 0.06495 .
foundational_course  1   50.6   50.57  0.3469 0.56029
Residuals          30 4373.3  145.78
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Model 3 F-test:
Analysis of Variance Table

Response: message_3_response_time
                   Df  Sum Sq Mean Sq F value  Pr(>F)
message_3           1  204.07 204.073  3.0529 0.09018 .
foundational_course  1  171.29 171.292  2.5625 0.11925
Residuals          32 2139.04  66.845
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Model 4 F-test:
Analysis of Variance Table

Response: message_4_response_time
                   Df Sum Sq Mean Sq F value Pr(>F)
message_4           1   31.3  31.271  0.1795 0.6750
foundational_course  1   34.0  33.963  0.1949 0.6621
Residuals          29 5053.4 174.257
```

Heterogeneous Treatment Effects (HTE) were explored by including interaction terms between the message timing (treatment/control) and whether the instructor teaches a foundational course. This interaction term was added to see if the impact of message timing on response time varies for instructors based on their teaching status. The analysis shows that the interaction terms in all messages (Message 1 through Message 4) were not statistically significant. This implies that the impact of message timing on response time does not change between instructors who teach foundational courses and those who do not, meaning that there is no HTE based on foundational course status.

```
Call:
lm(formula = message_1_response_time ~ message_1 * foundational_course,
    data = data_message_1)

Residuals:
    Min      1Q  Median      3Q     Max
-9.8485 -3.5222 -0.3652  2.8682 14.6444

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                   10.1818     1.9130   5.322 7.13e-06 ***
message_1                     -5.7273     2.7054  -2.117   0.0419 *
foundational_course           -0.5318     3.2201  -0.165   0.8698
message_1:foundational_course -0.1672     4.3014  -0.039   0.9692
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.345 on 33 degrees of freedom
Multiple R-squared:  0.1934,    Adjusted R-squared:   0.12
F-statistic: 2.637 on 3 and 33 DF,  p-value: 0.06594
```

```
Call:
lm(formula = message_2_response_time ~ message_2 * foundational_course,
    data = data_message_2)

Residuals:
    Min      1Q  Median      3Q     Max
-17.098  -7.817  -3.279   6.006  22.802

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                    17.165      4.340   3.955 0.000452 ***
message_2                      -7.787      5.704  -1.365 0.182705
foundational_course            -1.898      6.354  -0.299 0.767290
message_2:foundational_course  -1.151      8.695  -0.132 0.895620
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12.28 on 29 degrees of freedom
Multiple R-squared:  0.1186,    Adjusted R-squared:  0.02746
F-statistic: 1.301 on 3 and 29 DF,  p-value: 0.2929
```

```
Call:
lm(formula = message_3_response_time ~ message_3 * foundational_course,
    data = data_message_3)

Residuals:
    Min      1Q  Median      3Q     Max
-12.111  -5.693  -1.748   3.657  33.193

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                   7.8074     2.7688   2.820   0.0083 **
message_3                    -5.0762     4.0362  -1.258   0.2179
foundational_course           4.3204     3.9157   1.103   0.2784
message_3:foundational_course 0.2225     5.6235   0.040   0.9687
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.306 on 31 degrees of freedom
Multiple R-squared:  0.1493,    Adjusted R-squared:  0.06701
F-statistic: 1.814 on 3 and 31 DF,  p-value: 0.1652
```

```
Call:
lm(formula = message_4_response_time ~ message_4 * foundational_course,
    data = data_message_4)

Residuals:
    Min      1Q  Median      3Q     Max
-13.615 -10.046  -4.250   8.213  32.935

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                   9.5771     4.7422   2.020   0.0531 .
message_4                     0.5646     6.7065   0.084   0.9335
foundational_course           0.6479     6.7065   0.097   0.9237
message_4:foundational_course 2.8250     9.4845   0.298   0.7680
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13.41 on 28 degrees of freedom
Multiple R-squared:  0.01586,   Adjusted R-squared:  -0.08958
F-statistic: 0.1504 on 3 and 28 DF,  p-value: 0.9285
```

It is essential to note that the high standard errors in all models are probably due to the small sample size combined with many unobserved confounders and possible spillover effects. Overall, it looks like there are some patterns in response times based on message type, but with the high uncertainty, confounders, and not great explanatory power of our models, we could benefit from additional research - maybe simply starting with more data.

Ethical Considerations

The design of this study posed some ethical dilemmas. First and foremost, this social science field experiment was performed on human subjects without obtaining pre-approval from an institutional review board (IRB). This experiment was required for the UC Berkeley MIDS course on Experiments and Causal Inference. As such, the experimental design was pre-approved by the course instructor, an expert in social science experimental design. Furthermore, this course's short (14-week) timeline makes obtaining IRB approval and completing the experiment in time impractical.

Other ethical aspects of this research can be analyzed in the framework of the Belmont report [3]. According to this report, ethical research considerations on human subjects should address three categories: respect for persons, beneficence, and justice.

Respect for Persons/Autonomy: This study did not consent subjects before administering treatment and involved a degree of deception regarding the message's intent (the investigators did not always need the syllabus, life advice. etc.). To address this concern, the messages sent out were thoroughly vetted and discussed with the course instructor and were crafted to be very respectful of the participant. Furthermore, while the initial message was sent with consent, the attempt to obtain the more sensitive demographic variables was made after explaining the experiment to the participants and allowing them to opt into this part of the experiment by volunteering this information. Sending out a "warning email" to all MIDS instructors about the potential for such an experiment during the fall semester would have reduced the concern over respecting subject autonomy. However, the investigators believe this would have a high risk of compromising the study design and validity of its results.

Beneficence: In this study, the instructors did not directly benefit from receiving the treatment. The potential harm to the instructors, however, was also low. All the study participants were UC Berkeley instructors actively teaching during the study period. As such, these instructors are responsible for aiding the education of MIDS participants, including the investigators of this study. Furthermore, receiving four Slack messages over three weeks was felt to be an acceptable cognitive load on the participants. Additionally, the subjects were compensated for their participation with a gift card at the end of the study.

Justice: This study's randomization design made subject discrimination unlikely. Although some instructors may have been disadvantaged in their response time due to residing in different time zones, all instructors in the MIDS program expect to be available for students during Pacific time hours.

Limitations

While the study provides valuable insights into communication patterns within the MIDS program, it's crucial to acknowledge and address several noteworthy limitations that could impact the interpretation and generalizability of the findings.

Confounders: Firstly, the study did not account for a range of potentially influential unobserved confounders, such as marital status, age, workload, and residential time zone of the instructors. Efforts to collect these variables post-treatment were met with a low response rate of only 20%, hindering any meaningful inclusion in our analysis.

Spillover: Additionally, a spillover effect was observed among the participants. Some instructors who were part of the study sought assistance from their peers in the program, potentially affecting the experiment's outcomes. This type of interaction is often unavoidable in a close-knit community like MIDS, but it may have led to an underestimation of the average treatment effect. Another challenge was the broad time frame defined for the treatment group. We considered any message sent between 9 AM and 5 PM as 'during work hours,' which didn't differentiate between early morning and late afternoon responses. Although we aimed to narrow this window to between 11 AM and 1 PM for the treatment and 6 PM to 8 PM for the control, this was not consistently achievable.

Lack of Demographic Data: Furthermore, there was significant difficulty in obtaining demographic data. Respecting participant autonomy, we refrained from collecting sensitive demographic information without explicit consent. The post-treatment survey, unfortunately, had a low response rate, and due to time constraints, further data collection efforts were not feasible. The study's sample size, both in terms of participants and messages, was relatively small. This limitation was a contributing factor to the significant standard errors observed in the study.

Variability in Messages: Lastly, the four different messages sent to participants varied slightly to accommodate familiarity levels between the investigators and some instructors. Modifying the messages was essential to preserve the integrity of the study and prevent participants from suspecting their involvement in an experiment.

While these limitations present challenges, they also offer directions for future research. Understanding and addressing these aspects can pave the way for more refined studies in the field of digital communication in academic settings.

Conclusions

Overall, this was a fascinating experiment to run and very applicable to our day-to-day life as students. Our findings suggest that to reduce response time, students should aim to send their message during working hours, mainly if it involves a simple task for the instructor. However, our findings should be interpreted carefully since our experiment was underpowered, and working-hour sending times did not significantly reduce response rates for other message types. To answer our study question more definitively, the experiment needs to be repeated with a larger sample size and the collection of more demographic data.

References

[1] Aranda, J. H., & Baig, S. (2018). Toward "JOMO": The joy of missing out and the freedom of disconnecting. In Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services (pp. 1-8).

[2] EZ Texting. (2019). 2019 Text Communications Report. Retrieved December 4, 2023, from https://www.eztexting.com/press/2019-text-communications-report

[3] National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979). Ethical principles and guidelines for the protection of human subjects of research. The Belmont Report.