Computer Science
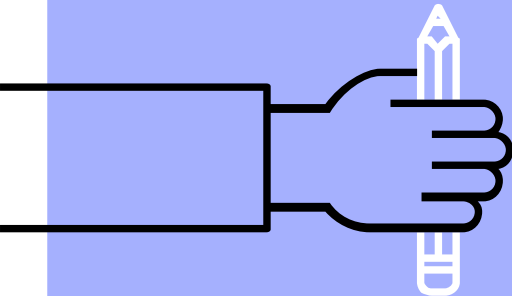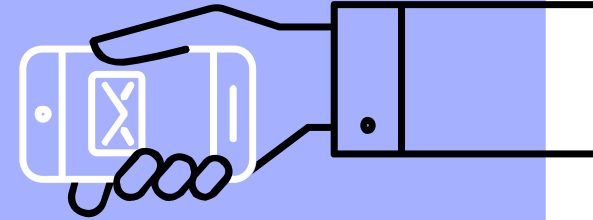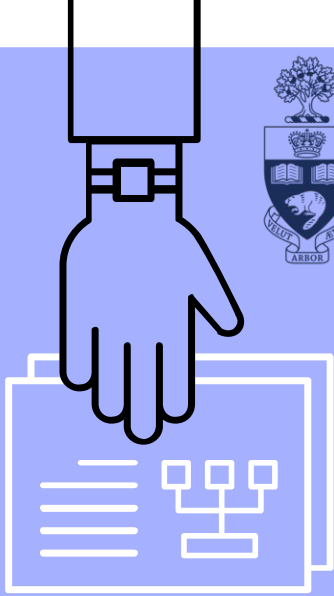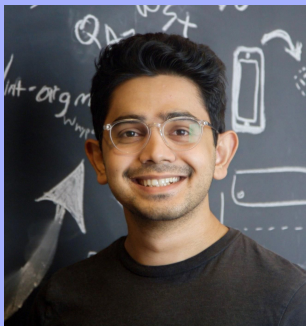UNIVERSITY OF TORONTO

SIGCSE '23 Workshop on Designing, Deploying, and Analyzing Adaptive Educational Field Experiments

Intelligent Adaptive Intervention
(IAI) Research Lab

https://intadaptint.org/
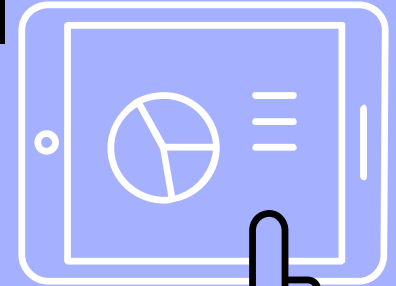
Mohi Reza
PhD Student
mohireza@cs.toronto.edu

Angela Zavaleta
PhD Student
angelazb@cs.toronto.edu

Tong Li
PhD Student
tongli@cs.toronto.edu

Nathan Laundry
Incoming PhD Student
nlaundry@uoguelph.cs

Ilya Musabirov
PhD Student
imusabirov@cs.toronto.edu

Joseph Jay Williams
Assistant Professor
williams@cs.toronto.edu

Michael Liut
Assistant Professor
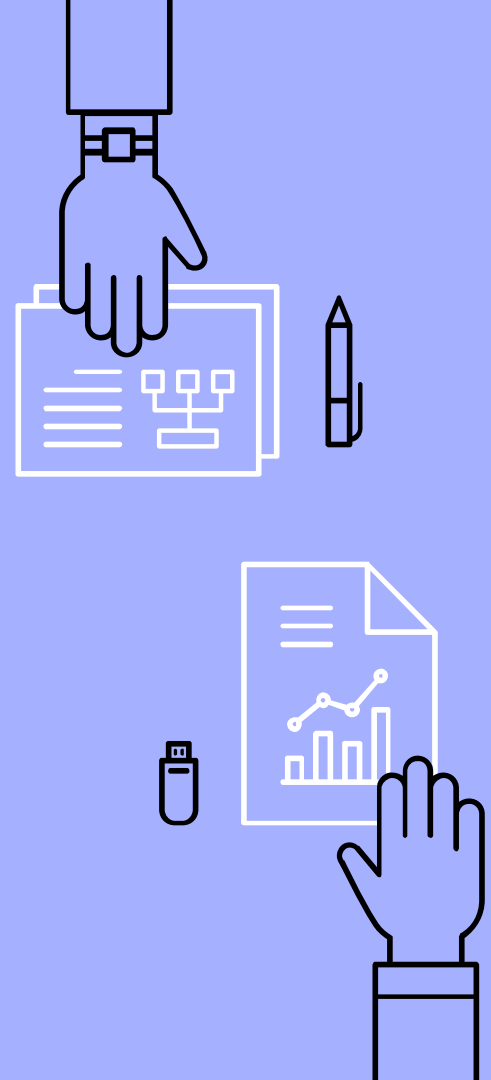liut@cs.toronto.edu

University of Toronto
Computer Science

tiny.cc/abhandout

Introduction

➔ How can we use A/B testing in education?

➔ How can we (as educators) benefit from A/B testing?

Workshop Structure

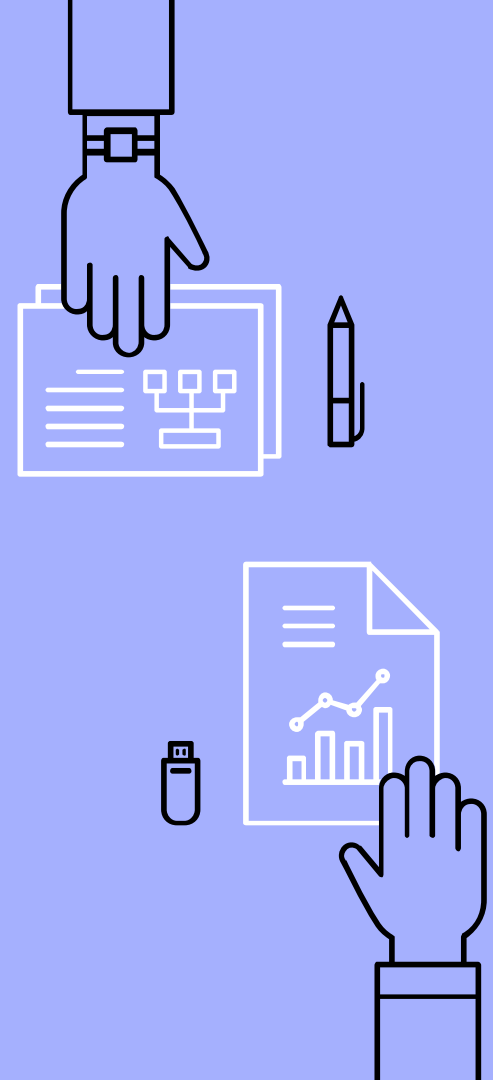This workshop will be comprised of four parts:

1. Pre-Workshop Survey
2. Designing Alternatives to A/B test
3. Data Analysis Walk-through
4. Post-Workshop Survey

Collecting some info!

For us to get started, we need you to complete a brief survey:
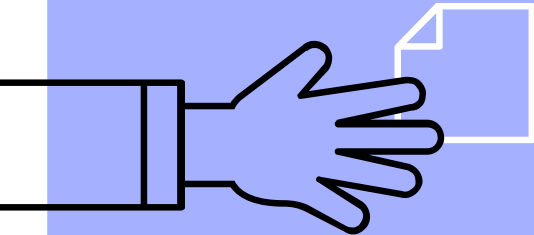
https://forms.gle/eKfnqpqaaCehF1AR7

Part 1
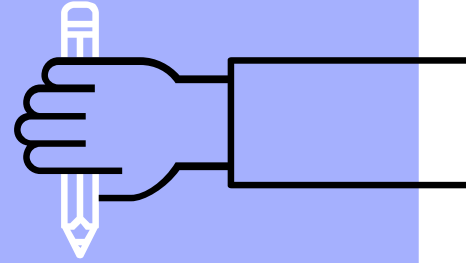
Experiment-Inspired Design

What if we take a broader view of what it means to run an education experiment?

What if instead, we used experimentation as a means educational designers to better think about the **design space** that they are working in?

"Traditional Experiment"

Encapsulates existing practices surrounding experimentation where the goal is often restricted to either **accepting or rejecting some pre-determined and well-defined hypothesis.**

"Experiment-Inspired Design"

Seeing experimentation as a means for **exploring the complex design space of generating and improving instructional content**, as a thinking tool for systematic design.

A/B Comparisons

A   **B**   **+** ⋯ **N**   Continually add conditions

50%   50%
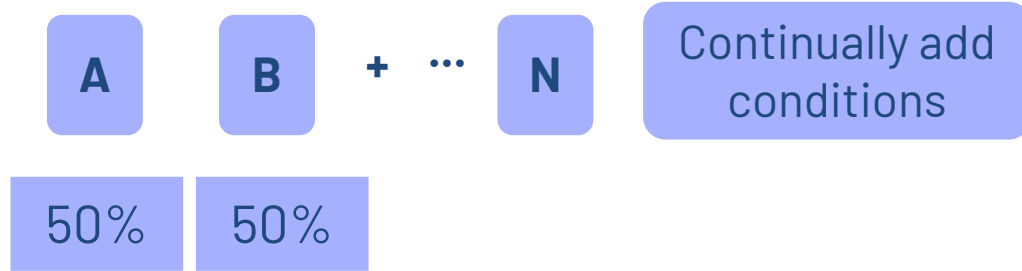
Factorial Design

Factors:

A   **B**   **+** ⋯ **N**   Continually add conditions

Values:

A1	B1	N1
A2	B2	N2
X	**X** ⋯	
⋮	⋮	⋮
AX	BY	NZ
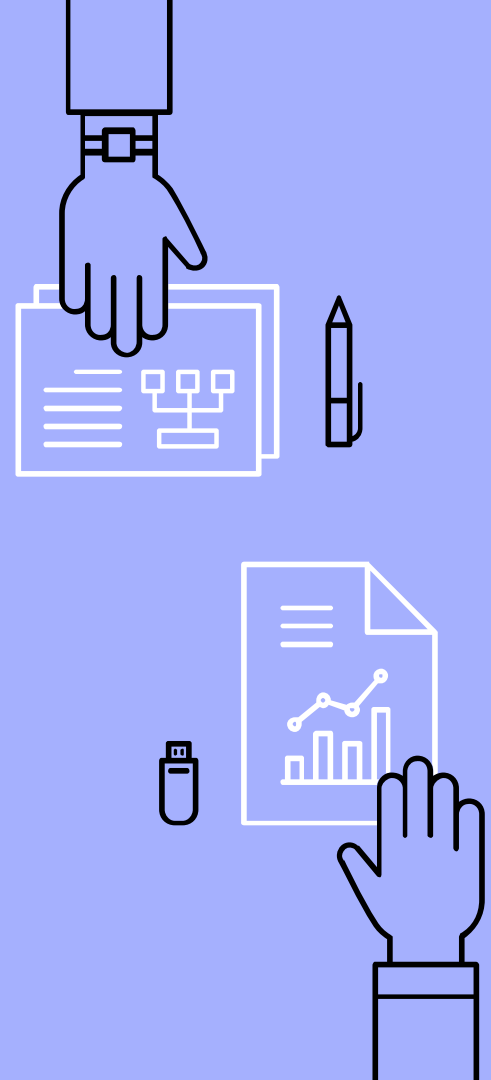
Real-World Examples
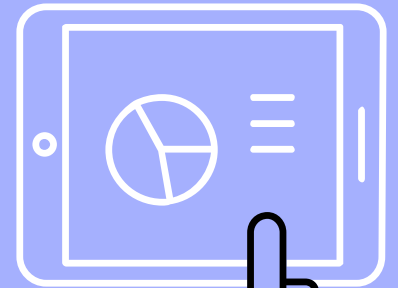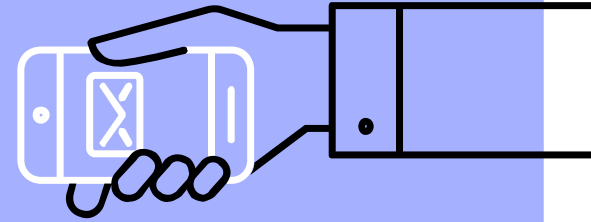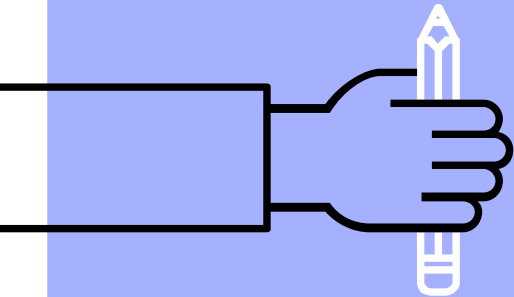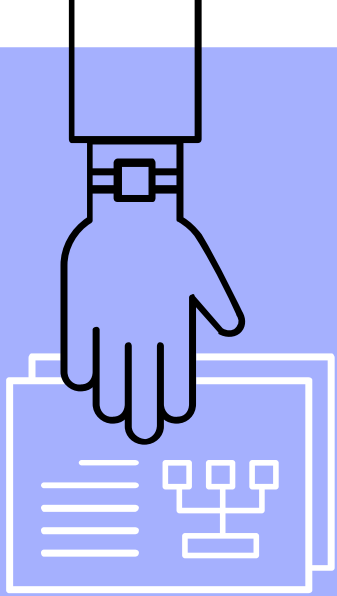
Real World Examples

Emails

"How do we get students to start their homework earlier?"

Homework Drop-Downs

"How do we motivate students to practice additional problems?"
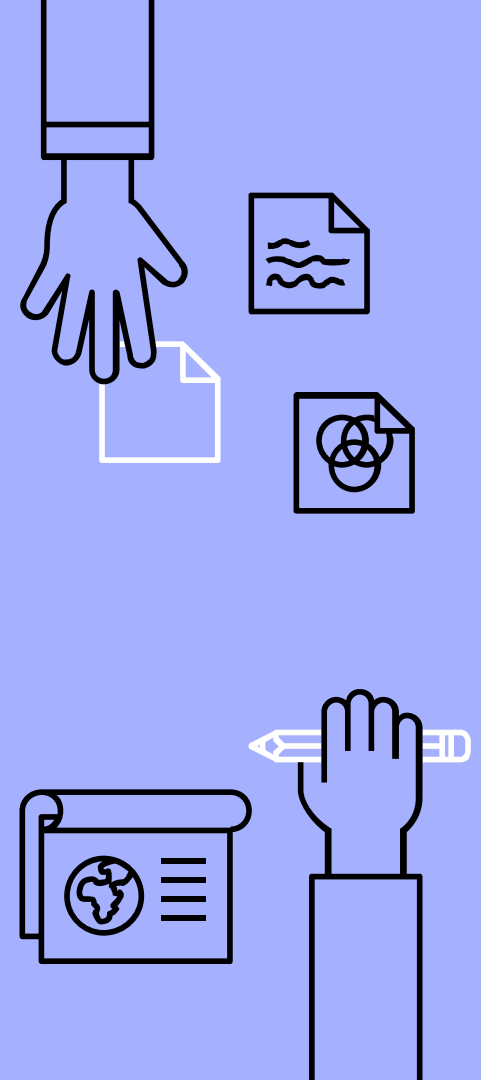
Emails

In this example, our goal is to test different ways we can **increase students' engagement with prompting emails that remind and provide advice on how to start their online homework early.**

We can A/B Test Different Ideas that we **think** could make our emails better using a Factorial Design

Email Factorial Design

Subject Line ✖ Information ✖ Sender

Prompt

| *When will you next work on your homework? Can you start earlier than last week?* |

No Information

| - |

Instructor

Reminder

| *Remember to start your homework early and finish before Friday at 5 pm* |

Information

| *We have found that students who start earlier tend to learn more, since if they get stuck, they have time to figure it out.* |

Teaching Assistants

Rogers et al., 2015

You, 2015

Goetz et al., 2021

Clarke et al., 2018

Email

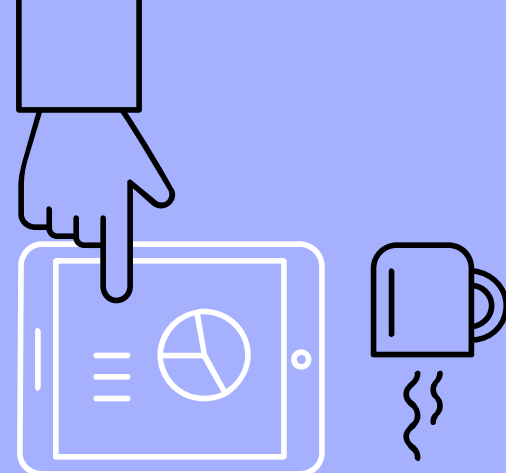**SLPrompt:** When will you next work on your homework? Can you start earlier than last week?
SLReminder: Remember to start your homework early and finish before Friday at 5 pm

Hi {{Name}},

The homework is due next Friday at 5pm. Please take a moment to think about the following prompts:

When will you do this week's homework? How can you start earlier than last week?

Information (include / not include): We have found that students who start earlier tend to learn more, since if they get stuck, they have time to figure it out.

Try to think of things that will help you follow through, like adding a time to your calendar.

Feel free to reply back and let us know how you plan to get an early start on this week's homework. Click here to access the homework.

Best regards,

{SenderInstructor/SenderTA}
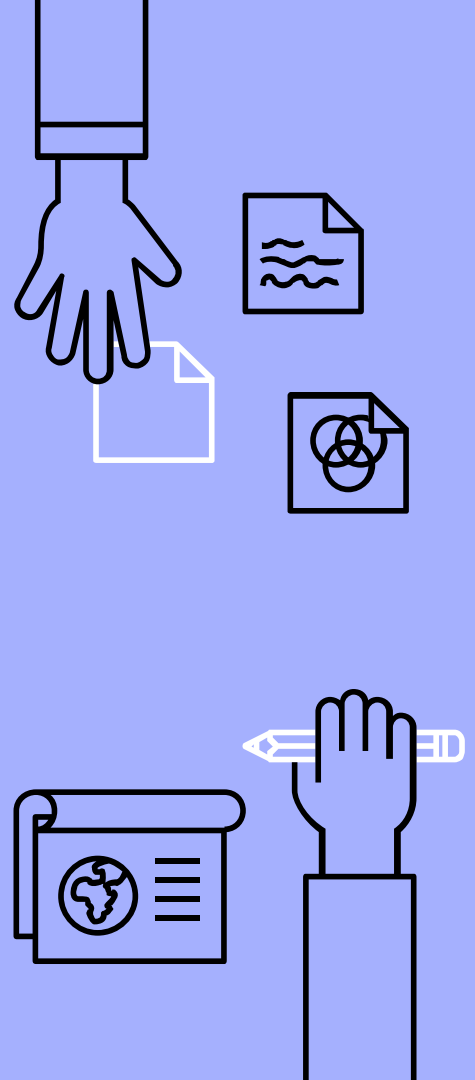
Homework Drop-Down

In this example, our goal is to test different prompts as a way to **increase students' engagement with optional tasks that appear after students complete an online homework problem.**

(1/4) Students attempt an online homework problem

aliasing ✔

After this code has been executed, what value does `list1[-1]` refer to?

```
list1 = [1, 2, 3]
list2 = list1
list2.append(4)
```

- [ ] 1
- [ ] 2
- [ ] 3
- [x] 4

History   Submit

(2/4)They see a randomized drop-down

Procrastination and bad time management are common problems stopping students from achieving great results

Solving a harder problem will help you to:

- train your brain to become better at programming problems
- become better at applying programming to various real-life contexts

Solve a harder problem
Skip

We can A/B Different Ideas for Prompts inside the Dropdowns using a Factorial Design

Drawing from Theory

An advantage of factorial experiments is that it can allow us to concurrently investigate multiple ideas and the connections between them. For example, based on prior research, we may be interested in grounding our prompts on two different theories:

- **Growth mindset** (the idea that one can grow their abilities through effort (Dweck, 2006)**:** train your brain to become better at programming problems
- **Mental Contrasting** (thinking about obstacles after imagining a positive future (Oettingen 2009)**:** Procrastination & bad time management are common problems stopping students from achieving great results.
- **Control**: Solving [a harder/another] problem [right now/-] will help you.

Mental Contrasting NO		Mental Contrasting YES	
Growth **M**indset NO	**G**rowth **M**indset YES	**G**rowth **M**indset NO	**G**rowth **M**indset YES
H&T: Solving [a harder/another] problem [right now/-] will help you	**H&T:** Solving [a harder/another] problem [right now/-] will help you **GM:** train your brain to become better at programming problems	**MC:** Bad time management & procrastination are common problems stopping students from achieving great results.	**MC:** Bad time management & procrastination are common problems stopping students from achieving great results. **H&T:** Solving [a harder/another] problem [right now/-] will help you **GM:** train your brain to become better at programming problems

<u>Note</u>: Hardness & Time (**H&T**) vary independent of this table

(3/4) They can click to see a optional harder problem

Assuming s refers to 'racecar', select the expression(s) that produce 'ace':

- s[1:3]
- s[1:len(s) - len(s) // 2]
- s[5:2:-1]
- s[1:len(s) // 2 + 1]
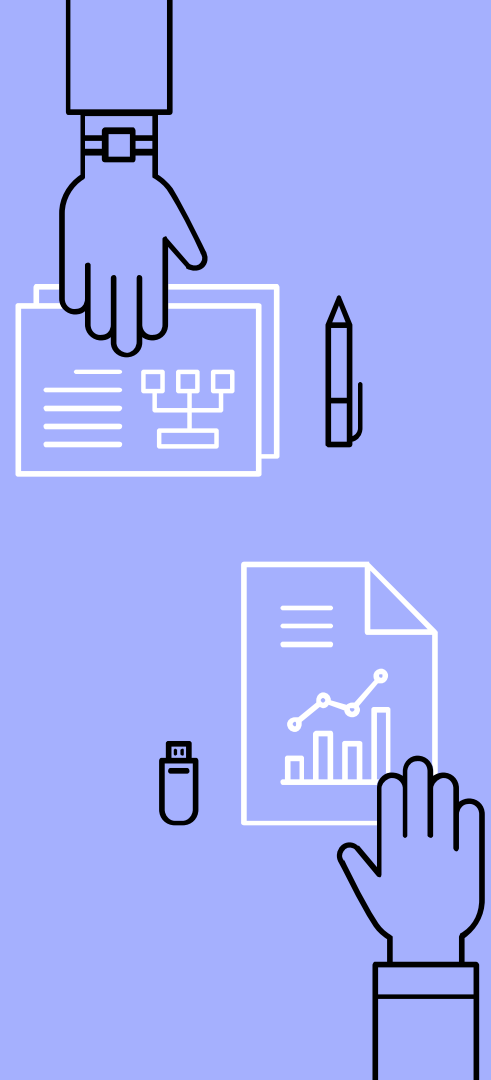- s[1:len(s) // 2]

››

(4/4) They can view the solutions afterwards

Task:
Assuming s refers to 'racecar', select the expression(s) that produce 'ace'
Correct answers are:

- s[1:len(s) – len(s) // 2]
- s[1:len(s) // 2 + 1]
- s[5:2:-1]

Read the paper at

tiny.cc/expdesignpaper

Experimenting with Experimentation: Rethinking The Role of Experimentation in Educational Design

Mohi Reza
mohireza@cs.toronto.edu
University of Toronto
Toronto, Ontario, Canada

Akmar Chowdhury
University of Toronto
Toronto, Ontario, Canada
akmar_ehsan@hotmail.com

Aidan Li
University of Toronto
Toronto, Ontario, Canada
aidan.mr.li@gmail.com

Mahathi Gandhamaneni
University of Toronto
Toronto, Ontario, Canada
mahathi.gandhamaneni@mail.utoronto.ca

Joseph Jay Williams
University of Toronto
Toronto, Ontario, Canada
josephjaywilliams@gmail.com

STRESS INTERVENTION
Being stressed during an exam can actually help you do better.

[A] Text on research-based encouragement
[B] Text on how to positively respond to stress
[C] Video on how to positively respond to stress
[D] Link to research paper on reappraising stress
[E] Prompt to reflect on how to use this information
[F] Prompt to write a helpful message as a reminder

OTHER ACTIVITIES

Core Message Shown to all students in the intervention group

Additional Content to A-B Test To reinforce the core message and test different content design ideas, we included a random subset of these six variants in addition to the core message

(a)

MCII INTERVENTION
[A] Baseline Planning Prompt   [B] WOOP Activity

Timeframe
[A] 24 hrs [B] 1 week [C] 1 month

Mid Activity Questions

Wish Format
[A] If-Then [B] Action-Time-Pace

Post Activity Questions

OTHER ACTIVITIES

Core Activity 50-50 random split between baseline and WOOP.

Variations to A-B Test To explore the design space of what a WOOP activity could look like, we A-B tested different timeframes and formats for expressing wishes.
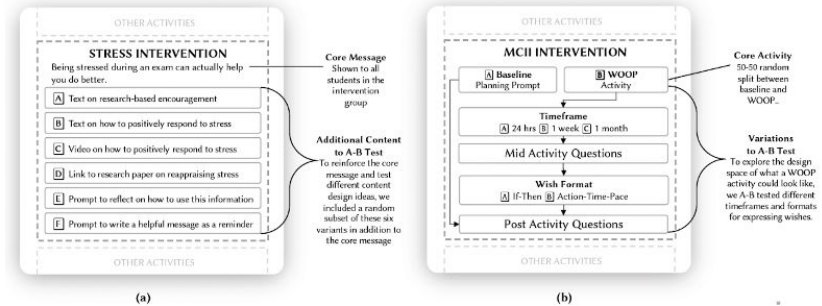
(b)

Figure 1: Summary of Intervention Design for the Stress Reappraisal and MCII Activities

ABSTRACT

What if we take a broader view of what it means to run an education experiment? In this paper, we explore opportunities that arise when we think beyond the commonly-held notion that the purpose of an experiment is to either accept or reject a pre-defined hypothesis, and instead, reconsider experimentation as a means to explore the complex design space of creating and improving instructional content. This is an approach we call *experiment-inspired design*. Then, to operationalize these ideas in a real-world experimentation venue, we investigate the implications of running a sequence of interventions teaching first-year students "meta-skills": transferable skills applicable to multiple areas of their lives, such as planning, and managing stress. Finally, using two examples as case studies for

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI)**; • **Applied computing** → **Education**;

KEYWORDS

A-B Testing, Education, Experimentation, Meta-Skills, Design Thinking, HCI

1 INTRODUCTION

Often, randomized experimentation is used in education research to evaluate an intervention, or test a specific hypothesis about teaching or learning. What if we take a creative perspective on the *goals* and *uses* of experiments than what is typically presented
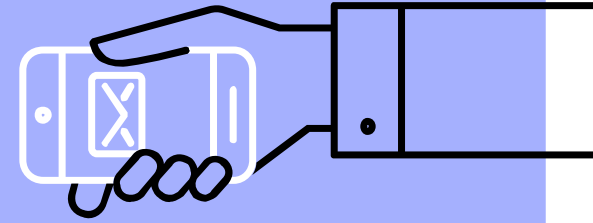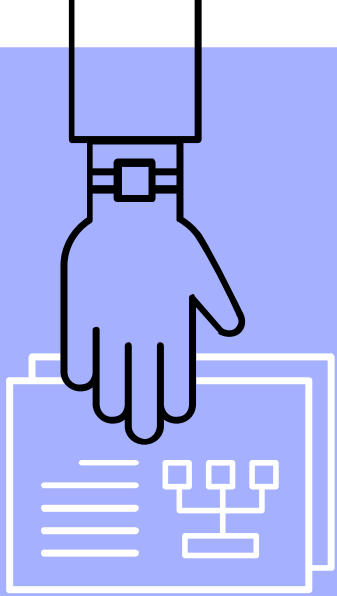
Exercise!
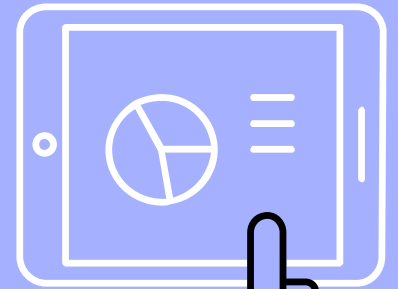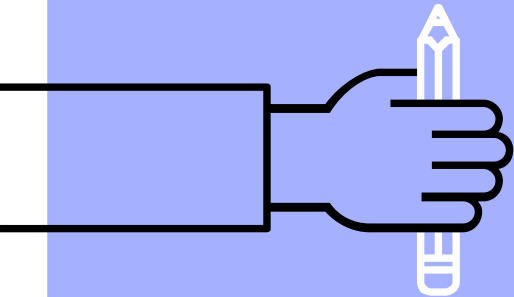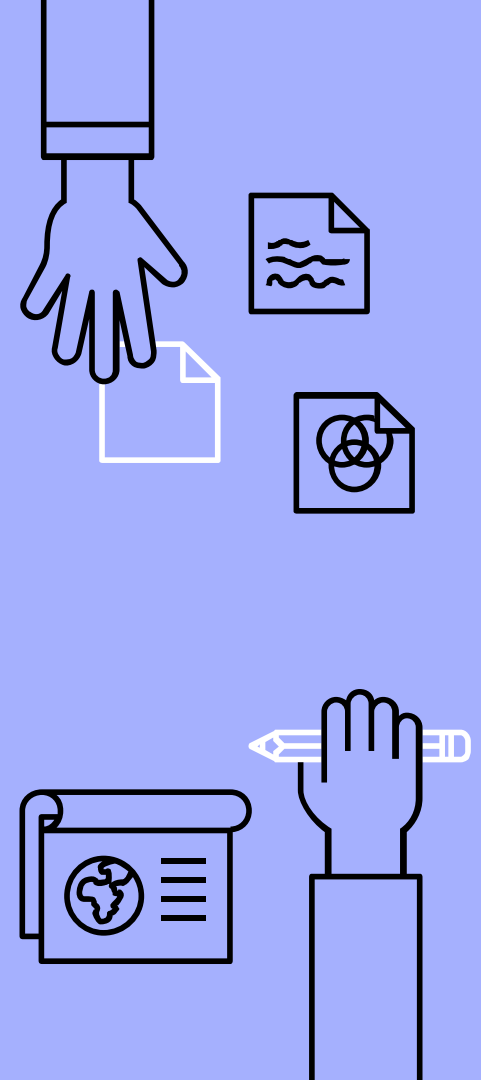
Splitting into Teams!

Emails

1. Person 1
2. Person 2
3. Person 3
4. Person 4
5. Person 5

Leader: Angela

Homework Drop-Downs

1. Person 6
2. Person 7
3. Person 8
4. Person 9
5. Person 10

Leader: Mohi

Emails

1. Skyler
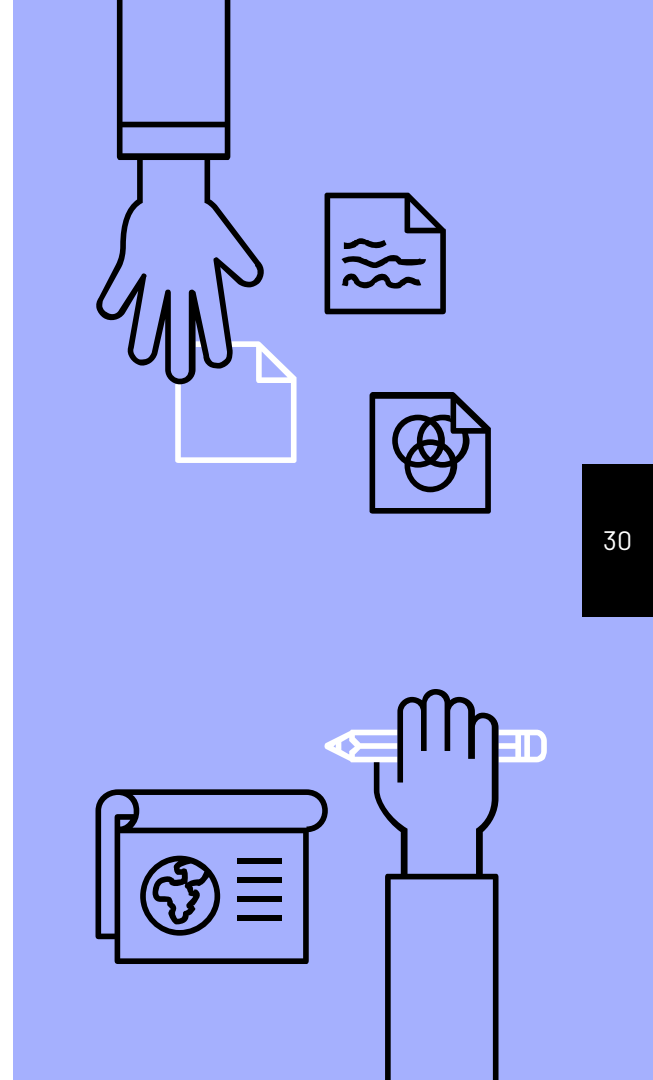2. Bruce
3. Steve
4. Isabella
5. Harry

Leader: Angela

Homework Drop-Downs

1. Vera
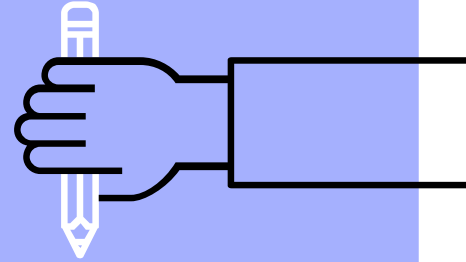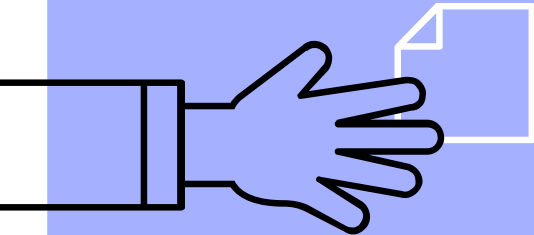2. Daisuke
3. Marko
4. Jessica
5. Leonardo

Leader: Mohi

tiny.cc/abemail

tiny.cc/abdropdown

Part 2

Analysis and Adaptation of Quantitative Data
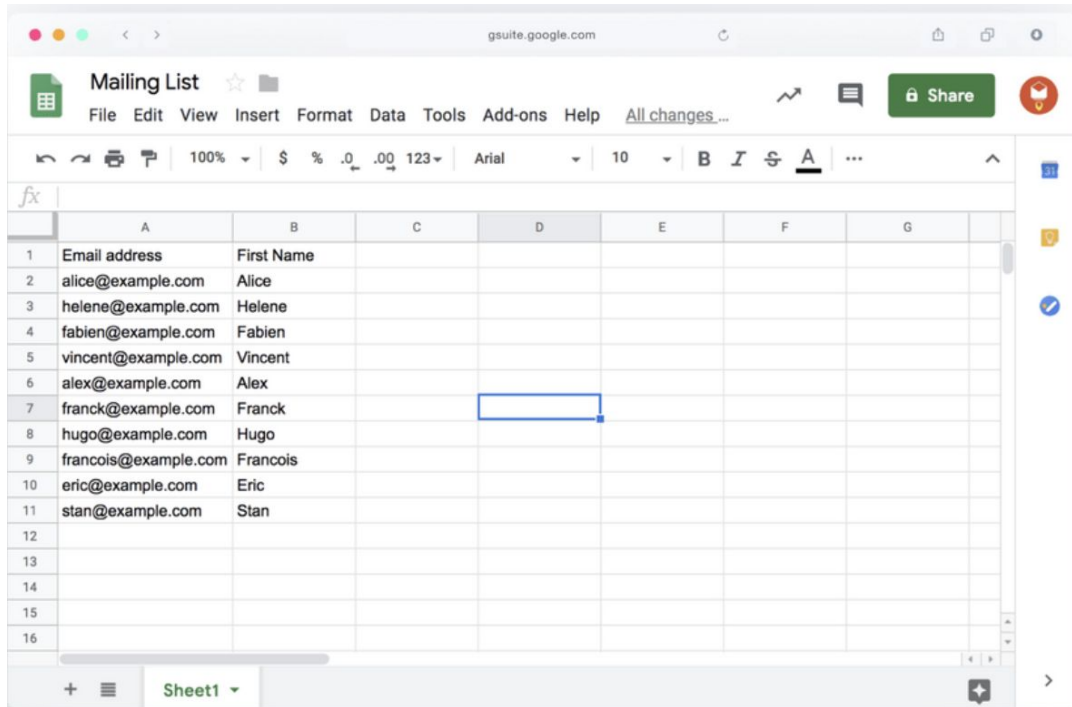
Deployment of A/B Comparisons

<u>Yet Another Mail Merge (YAMM)!</u>

- YAMM is a Google Chrome extension.
- Allows users to send personalized emails while tracking engagement metrics:
  - whether an email was opened or not
  - email open time
  - whether embedded links were clicked or not
- Allows the recipient to unsubscribe from an email subscription.
- Free!
  - ➤ capacity of ~50 emails a day (~400 with premium)
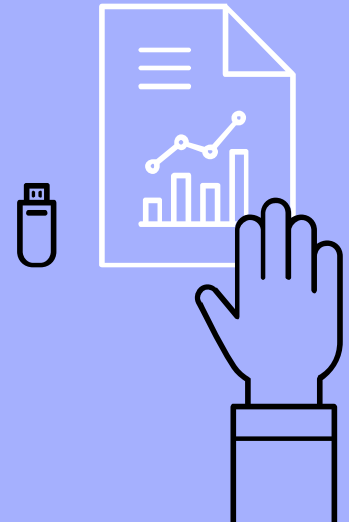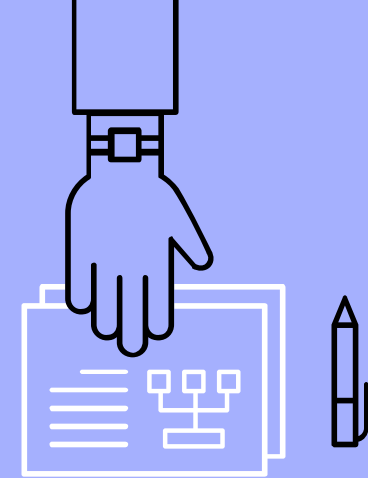
Create a Mail List!

Create a Draft Email!

Send the Email!

Survey Tool: Qualtrics!

- Qualtrics is a very popular tool for collecting quantitative and qualitative feedback.

XM FOR
Customer Frontlines

Software to deliver better experiences in contact centers and across every digital and physical touchpoint.

Digital   Care   Locations

XM FOR
People Teams

Software to engage teams, improve manager effectiveness, and make informed people and business decisions.

Engage   Lifecycle   Analytics

XM FOR
Strategy & Insights

Software to design products people love, increase market share, and accelerate growth.

Research   UX   Brand

Fernando Ruiz
USA, Illinois, Chicago
3:09

Phone number   (801) 374-6682
Email   fernandorchl@gmail.com
LTV
Member Since

Remember to
Show empathy

73%
TEAM ENGAGEMENT

57%   ACTION NEEDS
Set clear career goals
▼ 12

Brand Momentum Score   76
SEE DRIVERS

Link Tracking!

Edit end of survey

▼ **Messaging**

End of survey message

Redirect to URL

Website URL

https://www.loom.com/share/3592d0

Link Tracking!

Edit end of survey

▾ **Messaging**

End of survey message

Redirect to URL ⌄

Website URL

https://www.loom.com/share/3592d0

▸ | ED | **Set Embedded Data:**
| id | Value will be set from Panel or URL.   Set a Value Now

Link Tracking!

Set Embedded Data:

id Value will be set from Panel or URL. Set a Value Now

- Share the qualtrics survey link to your participants, and use a personal identifier to modify the links. For instance: https://.qualtrics.com/jfe/form/YYYYYY?id=user0 001 would be a link for user0001, and when this person clicks on it, it will get recorded on the Qualtrics logs.

Statistical Analysis of A/B Comparisons

➢ **Arm Mean and Standard Error of Arm Mean**
➢ **Wald (Chi-Squared) Test**
  ➢ https://docs.google.com/document/d/1PXpC2iDwjQ-kbUZNuDcM5X2sDxIBHFLGY0zef98LVuQ/edit#heading=h.2sovbybyxbfx

➢ **Interactive Data Activity!**
 ➢ **https://docs.google.com/spreadsheets/d/1-vjZdZP3Jg0WDvTM2kZN9F9dO7KyInaaOFOEx-ynb6c/edit#gid=0**

Conclusion!

✓ Experiment-Inspired Design
✓ Real World Examples
✓ Deployment of A/B Comparisons
✓ Data Collection & Analysis

Some Extra Resources!

- AXIS: [Generating Explanations at Scale with Learnersourcing and Machine Learning](#), by Williams et al.
- CHI'18: [Enhancing Online Problems Through Instructor-Centered Tools for Randomized Experiments](#), by Williams et al.
- KDD'07: [Practical Guide to Controlled Experiments on the Web: Listen to Your Customers not to the HiPPO](#), by Kohavi et al.
- WWW'14: [Designing and Deploying Online Field Experiments](#), by Bakshy et al.
- KDD'15: [From Infrastructure to Culture AB Testing Challenges in Social Networks](#), by Ya Xu et al.
- JEduTech: [Design-Based Research and Educational Technology: Rethinking Technology and the Research Agenda](#), by Amiel et al.
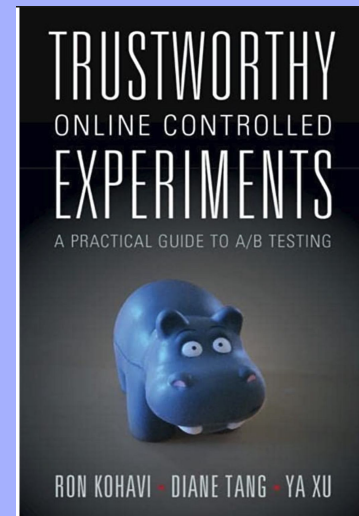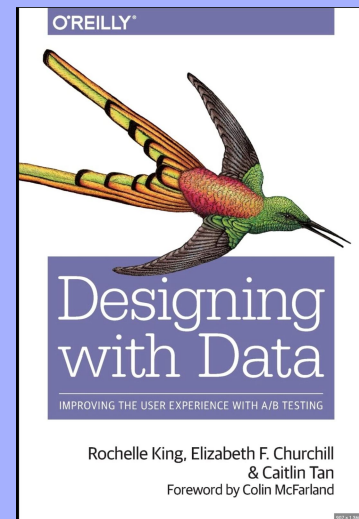
SIGCSE's Post-Survey

For you to provide feedback to SIGCSE:

https://www.surveymonkey.com/r/onlineworkshops