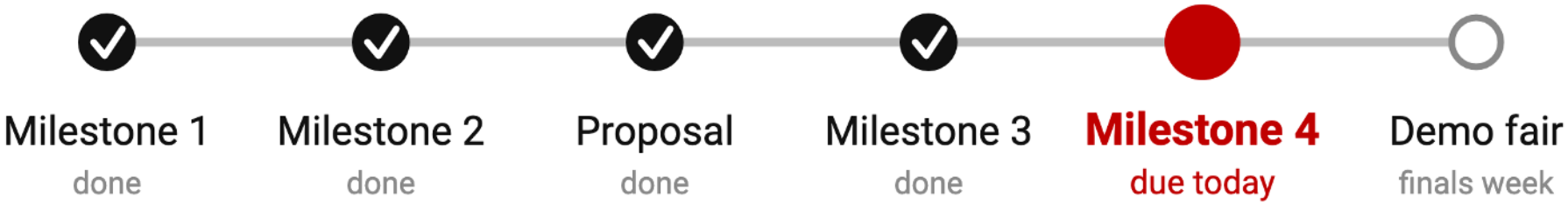
Human-AI Interaction, Lecture 19

# Project Studio

# Milestone 4: Decide what it does on its own

Due tonight at 9 pm



# Today: three rounds of testing




I call time at the end of each part.

# One round of testing

**test**

10 minutes

the host gives one task;
the visitor thinks aloud

**critique**

5 minutes

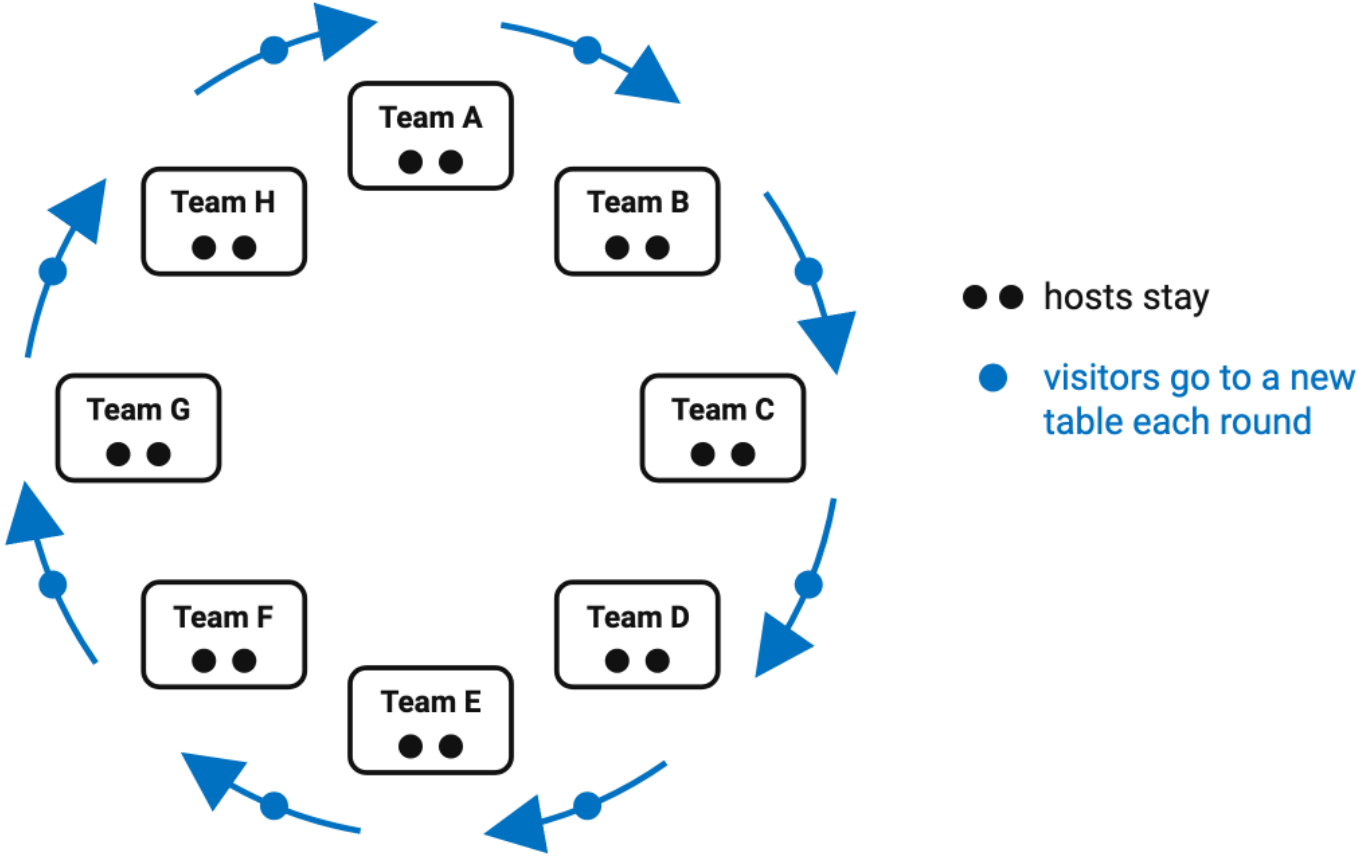
the visitor fills in a card
and talks it through

**move**

2 minutes

visitors go back;
the next ones head out

# Rotation between tables



# Thinking aloud

Definition: In a **thinking aloud test,** you ask test participants to use the system while continuously thinking out loud — that is, simply verbalizing their thoughts as they move through the user interface.

Jakob Nielsen, “Thinking Aloud: The #1 Usability Tool,” NN/g, 2012

If they go quiet:

**“What are you thinking?”**

**“What did you expect?”**

# Testing with classmates

Classmates can show whether people understand, steer and delegate to your tool.

Ask your own users whether they **want** it.

# Consent

1. Say what the session is for
2. Say what you will record, and who will see it
3. Say they can stop at any time

# The task card

A vague task

**“Try out our tool.”**

# The task card

A vague task

“Try out our tool.”

**A real task**

“You run the chess club. The meeting moved to Thursday. Get the reminder email ready to send to the members.”

Your task should include:

1. something the tool does **on its own**
2. a moment where it **asks first**
3. a **mistake you planted**

# The note sheet

**Test notes** Team _____ Round ___ Visitor _____

| Time | What the tool did | What the person did or said | Relied on it |
|---|---|---|---|
| *3:40* | *changed room 204 to 240 without asking* | *“is it in 240 now? OK, send it.”* | *too much: sent it without checking* |
| | | | |
| | | | |
| | | | |

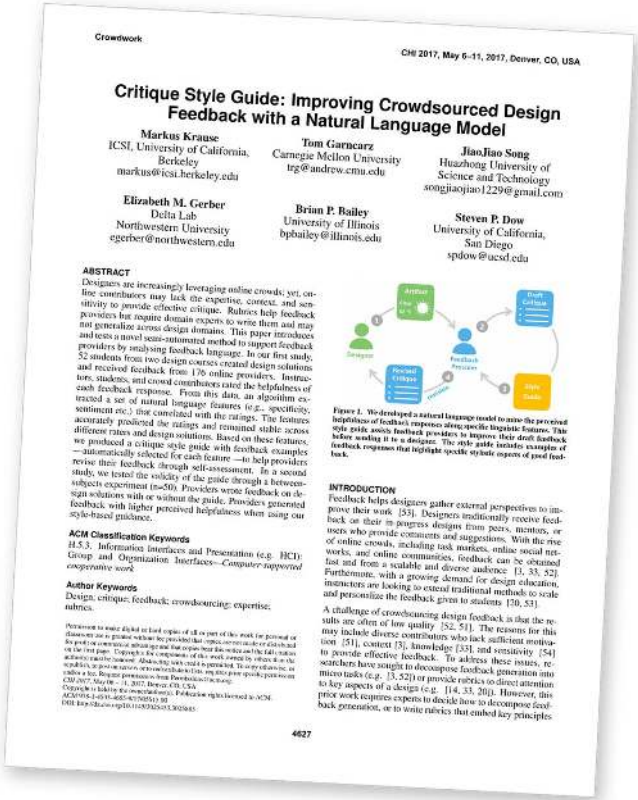
Relied on it: too much (kept it when it was wrong) · too little (redid it when it was right) · about right

# Questions after the task

1. What did the tool do on its own?
2. When it asked you first, was that the right moment?
3. Would you send this as it is? Why?

# Krause et al.: Critique Style Guide

Feedback that was specific and gave reasons tended to be rated **more helpful**.



Crowdwork

CHI 2017, May 6–11, 2017, Denver, CO, USA

**Critique Style Guide: Improving Crowdsourced Design Feedback with a Natural Language Model**


**Markus Krause**
ICSI, University of California, Berkeley
markus@icsi.berkeley.edu

**Tom Garncarz**
Carnegie Mellon University
trg@andrew.cmu.edu

**JiaoJiao Song**
Huazhong University of Science and Technology
songjiaojiao1229@gmail.com

**Elizabeth M. Gerber**
Delta Lab
Northwestern University
egerber@northwestern.edu

**Brian P. Bailey**
University of Illinois
bpbailey@illinois.edu

**Steven P. Dow**
University of California, San Diego
spdow@ucsd.edu



**ABSTRACT**

Designers are increasingly leveraging online crowds; yet, online contributors may lack the expertise, context, and sensitivity to provide effective critique. Rubrics help feedback providers but require domain experts to write them and may not generalize across design domains. This paper introduces and tests a novel semi-automated method to support feedback providers by analysing feedback language. In our first study, 52 students from two design courses created design solutions and received feedback from 176 online providers. Instructors, students, and crowd contributors rated the helpfulness of each feedback response. From this data, an algorithm extracted a set of natural language features (e.g., specificity, sentiment etc.) that correlated with the ratings. The features accurately predicted the ratings and remained stable across different raters and design solutions. Based on these features, we produced a critique style guide with feedback examples —automatically selected for each feature —to help providers revise their feedback through self-assessment. In a second study, we tested the validity of the guide through a between-subjects experiment (n=50). Providers wrote feedback on design solutions with or without the guide. Providers generated feedback with higher perceived helpfulness when using our style-based guidance.



**ACM Classification Keywords**

*H.5.3.* Information Interfaces and Presentation (e.g. HCI): Group and Organization Interfaces—*Computer-supported cooperative work*

**Author Keywords**

Design; critique; feedback; crowdsourcing; expertise; rubrics.



Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
*CHI 2017, May 06 – 11, 2017, Denver, CO, USA*
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-4655-9/17/05/$15.00
DOI http://dx.doi.org/10.1145/3025453.3025865





**Figure 1. We developed a natural language model to mine the perceived helpfulness of feedback responses along specific linguistic features. This style guide assists feedback providers to improve their draft feedback before sending it to a designer. The style guide includes examples of feedback responses that highlight specific stylistic aspects of good feedback.**

**INTRODUCTION**

Feedback helps designers gather external perspectives to improve their work [53]. Designers traditionally receive feedback on their in-progress designs from peers, mentors, or users who provide comments and suggestions. With the rise of online crowds, including task markets, online social networks, and online communities, feedback can be obtained fast and from a scalable and diverse audience [3, 33, 52]. Furthermore, with a growing demand for design education, instructors are looking to extend traditional methods to scale and personalize the feedback given to students [20, 53].

A challenge of crowdsourcing design feedback is that the results are often of low quality [52, 54]. The reasons for this may include diverse contributors who lack sufficient motivation [51], context [3], knowledge [33], and sensitivity [54] to provide effective feedback. To address these issues, researchers have sought to decompose feedback generation into micro tasks (e.g., [3, 52]) or provide rubrics to direct attention to key aspects of a design (e.g. [14, 33, 20]). However, this prior work requires experts to decide how to decompose feedback generation, or to write rubrics that embed key principles



Krause et al., “Critique Style Guide,” CHI 2017

# Four reminders from the guide

“Make sure your feedback is specific enough!”

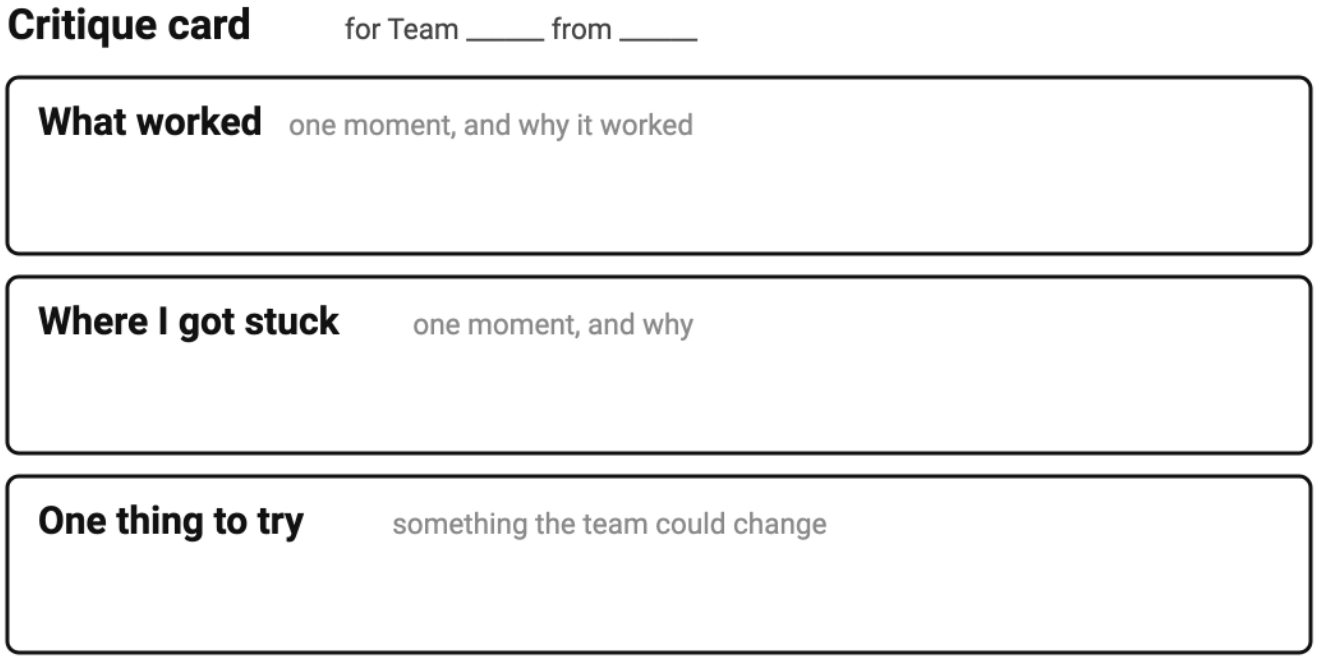
“Please make sure you explain your judgement!”

“Does your feedback suggest ways to improve the submission?”

“Make sure your feedback isn't too negative.”

Krause et al., CHI 2017, Table 4

# The critique card: three boxes




Each box is about **one moment** from the test.

# The critique card: three boxes

**Critique card** for Team _____ from _____

**What worked** one moment, and why it worked

When it asked before sending, it showed me who would get the email, so I knew what I was saying yes to.

**Where I got stuck** one moment, and why

It changed the room number on its own and I only noticed at the end, because nothing on the screen said it had changed anything.

**One thing to try** something the team could change

Mark what it changed since my last edit, and let me undo just that.

When it's your tool, **write it down** and say thank you.

# Team debrief: 8 minutes

1. Put each moment under **Understand**, **Steer** or **Delegate**
2. Pick **three changes** to make before the demo fair
3. Write down why you picked them

# Finals week: the demo fair




Photo: Omar David Sandoval Sida, Wikimedia Commons, CC BY-SA 4.0

# What to bring

1. A **working demo** on a charged laptop, and the charger
2. Your milestone 3 video
3. **One task** a visitor can try in two minutes

# How the fair runs

1. Half the team stays at the table, half walks around
2. **Switch halfway**
3. Each visitor tries **one task**, then you talk

Due two days after the fair

## **Final paper**

One per team, in the style of a CHI Late-Breaking Work

# Next: Lecture 20, AI as Normal Technology

Read before class:

- Narayanan & Kapoor, “AI as Normal Technology,” Part I, 2025
- Farrell et al., “Large AI models are cultural and social technologies,” 2025

**Bring the gulf map you made in Lecture 1.**