When to Give Feedback: Exploring Tradeoffs in the Timing of Design Feedback

Jane L. E*
je@ucsd.edu
University of California, San Diego
La Jolla, USA

Yu-Chun Grace Yen*
yyen@cs.nycu.edu.tw
National Yang Ming Chiao Tung University
Hsinchu, Taiwan

Isabelle Yan Pan
isabelle.yan.pan@gmail.com
University of California, San Diego
La Jolla, USA

Grace Lin
graceerya@gmail.com
University of California, San Diego
La Jolla, USA

Mingyi Li
mil011@ucsd.edu
University of California, San Diego
La Jolla, USA

Hyoungwook Jin
jinhw@kaist.ac.kr
KAIST
Daejeon, Republic of Korea

Mengyi Chen
mec005@ucsd.edu
University of California, San Diego
La Jolla, USA

Haijun Xia
haijunxia@ucsd.edu
University of California, San Diego
La Jolla, USA

Steven P. Dow
spdow@ucsd.edu
University of California, San Diego
La Jolla, California, USA

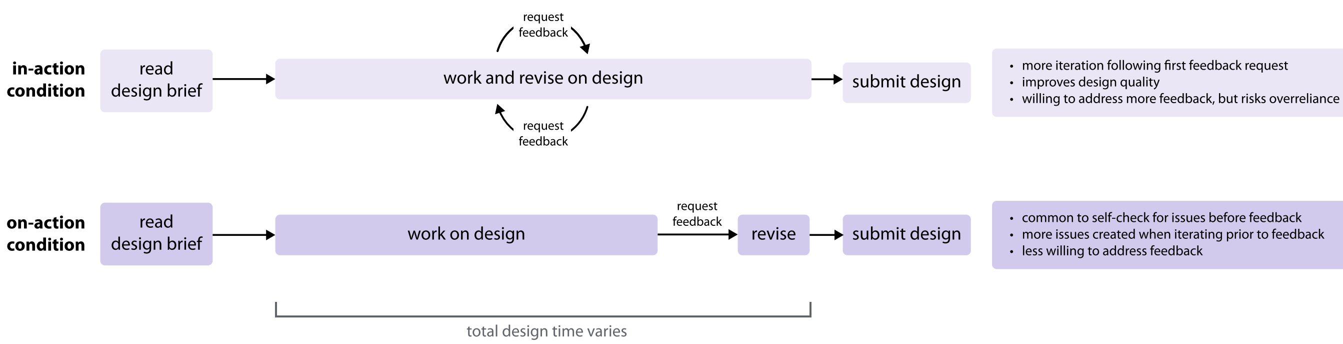


Figure 1: Inspired by Schön's distinction between reflection-in-action and reflection-on-action, this paper investigates two timing conditions for feedback within the context of a single design session: in-action feedback available *throughout* the design session (top) and the more commonly available on-action feedback on an completed design draft *after* the current design session (bottom). Here we diagram the flow of our two conditions (left) and summarize our main study findings (left).

ABSTRACT

Advances in AI have opened up the potential for creativity tools to computationally generate design feedback. In a future when designers can request feedback anytime on demand, how would the timing of these requests impact novices' creative learning processes? What are the tradeoffs of providing access to feedback *throughout* a design task (**in-action**) versus only providing feedback *after* (**on-action**)? We explored these questions through a Wizard-of-Oz study (N=20) using an interactive design probe, where participants could request feedback either throughout the design process or only after they complete a full draft. We found that in-action participants frequently request feedback, resulting in better improvements as indicated by a greater decrease in issues in their final design. However, we saw that in-action feedback can also risk users overly relying on feedback instead of engaging in more holistic self-evaluation. We discuss the implications of our insights on designing tools for creative feedback.

*Both authors contributed equally to this research.

C&C '24, June 23–26, 2024, Chicago, IL, USA

ACM ISBN 979-8-4007-0485-7/24/06
https://doi.org/10.1145/3635636.3656183

CCS CONCEPTS

• **Human-centered computing** → **Interactive systems and tools**.

KEYWORDS

visual design; feedback; creativity support tools; human-AI collaboration; empirical studies of design

ACM Reference Format:
Jane L. E, Yu-Chun Grace Yen, Isabelle Yan Pan, Grace Lin, Mingyi Li, Hyoungwook Jin, Mengyi Chen, Haijun Xia, and Steven P. Dow. 2024. When to Give Feedback: Exploring Tradeoffs in the Timing of Design Feedback. In *Creativity and Cognition (C&C '24), June 23–26, 2024, Chicago, IL, USA*. ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3635636.3656183

1 INTRODUCTION

Despite a wealth of educational resources on design (e.g., websites, books, and videos), novices often learn, not just by reading theory on design, but through getting practice and receiving expert feedback [2, 37, 93], which can encourage reflection and iteration [5, 107]. In formal design classes, students typically receive critiques on project work, which is often constrained by the size and availability of the teaching team [7, 23], limiting both the frequency and quantity of feedback. Researchers have worked on filling that gap by enabling novices to seek feedback from other sources, such as: peers [63], crowds [29, 71, 104, 106], or online communities [16, 39, 54].

A key challenge with exchanging feedback is that it often comes after a delay. Designers often need to wait to get input on their designs due scarcity of feedback providers [16, 79, 109] or simply taking the time to understand the intent and write thoughtful feedback [39, 77]. These challenges can preclude novices from even asking for feedback [39, 46, 54, 60] or compel them to "polish" their design before requesting feedback [74], which can increase fixation and limit exploration [6, 28, 30, 52].

To address situations where human feedback providers are not available, numerous research projects have explored the potential for integrating AI within a design tool. Creativity tools that envision interactions with AI often focus on generating design inspiration [19, 48], examples [62, 64], or variants on the existing design [84, 97, 98, 100]. A few tools have focused on creating personalized feedback that could help creators learn [89, 92]. For instance, Fischer et al. first explored embedding "critics" that could guide users towards their design goals for an interior design task [38]. Real-time feedback can prompt reflection before a designer gets too fixated on a single concept, but it can also be distracting and eventually discounted if poorly executed (much like users' reactions to Clippy [103]). Despite this push towards facilitating automated feedback, the research community still lacks solid empirical data on how designers—novices in particular—might engage with and learn from feedback depending on the timing. Our research not only explores how we might embed automated feedback within a design tool, but also how the timing of that feedback could impact novice designers engagement with feedback.

Schön described the importance of reflective dialogue between a designer and their work both during and after a design session [93]. The availability of feedback can influence the type of reflection: *reflection-on-action* arises from slower-paced feedback at discrete intervals, focused on work that has already been done (similar to getting feedback on homework from a teacher). *Reflection-in-action* happens when receiving early and frequent feedback that actively shapes an in-progress design (akin to getting grammatical feedback within a writing tool) [15, 41, 93]. Our research explores the following questions: If designers could request feedback from a design tool anytime on demand, at what timing and frequency would they request and how would it impact novices' creative learning processes? What are the tradeoffs of providing access to feedback *throughout* a design task (**in-action**) versus only providing *post-hoc* feedback (**on-action**)? While many projects have focused on the technology for enabling real-time feedback, our work seeks to understand what considerations are important in designing such tools, especially regarding creative learning.

To explore these questions, we developed a visual design tool—as a Wizard-of-Oz system [76]—that enables requesting feedback on demand. Our system aims to provide meaningful feedback (*high-quality*), that is available any time (*on-demand*) and personalized to the user's work and process (*context-aware*). Leveraging a hidden wizard allowed the users—who assumed interventions to be *computational*—to request feedback without the typical social stresses [51, 54, 57, 60, 74].

We explored the question of feedback timing through two studies. In a preliminary study, participants were asked to spend some time designing a poster in the tool. We gained initial insight into how and when users respond to in-action feedback and iterated on the design tool and protocols. Then, in a between-subjects experiment, 20 novice designers were randomly assigned to either the **in-action condition** (that allows feedback requests at any time during the design session) or the **on-action condition** (that gives a single feedback opportunity after a completed draft, simulating current practice). Experts in visual design served as the wizards in the study, generated personalized feedback for all participants throughout the design tasks using a blind-to-condition approach.

We found the in-action participants often requested feedback shortly after having a basic skeleton of a design and were more willing to address the feedback they received, resulting in more improvement in design quality during the design session, as indicated by a decrease in the total number of expert-identified issues. Since on-action participants were only allowed to request feedback one time (simulating a critique), they tended to wait until later to make this request. Of the feedback they were given, they were also less likely to address feedback resulting in significantly less change in their designs than their in-action counterparts.

We also report on the potential tradeoffs of feedback timing on participants' creative processes. While in-action feedback enables creators to reflect and evaluate their designs based on design principles, it also risks overly relying on feedback, focusing on small details, and giving less attention to their overall vision. We discuss the broader implications of our insights on the development of tools for computational feedback in creative domains.

Our research makes the following contributions:

- A Wizard-of-Oz **design probe** that offers preliminary interface designs for how visual design tools might embed high-quality, on-demand, context-aware, and computational feedback. The probe additionally includes the wizard interface for providing feedback and an administrative interface for managing the experiment.
- **Empirical evidence** from a between-subjects experiment (N=20) on feedback timing. We find that in-action feedback can lead to increased engagement and fewer design issues, while on-action feedback might allow more creative freedom and exploration.

2 RELATED WORK

We review related literature on feedback for learning, reflection-in and on-action, and tools for creativity support, and describe how they intersect with our questions about the timing of feedback.

2.1 Characteristics of Effective Feedback

Researchers have long studied the characteristics of effective feedback [18, 20, 22, 81, 109] and explored methods to encourage providers to generate feedback that satisfies these qualities [71, 79]. Learning in creative domains relies on iterating and reflecting on one's existing knowledge and creations. Sadler [89] describes how effective feedback for learning should encourage more reflection, which in turn facilitates an understanding of how to iterate. Sadler proposes that good feedback should be specific, justified, and actionable. We describe our approach to operationalize these characteristics in our "computational" feedback.

Many researchers have studied how important reflection is to being able to learn through experience and process-oriented methods such as experiential learning and deliberate practice [2, 37, 54, 56, 57, 69, 92]. Feedback can also act as a catalyst for promoting self-reflection during the creative process [37, 54], often scaffolded by experts [25, 91]. To further support the learning that happens within the process of "doing" [92], prior research has embedded feedback into design tools [38, 64, 80]. Taking inspiration from these tools and the characteristics of effective feedback, we created a visual design tool that enables high-quality "computational" feedback on demand, to better understand how novices interact with such feedback. By enabling real-time feedback, our research probe enables empirically investigating feedback timing and its impact on creative learning.

2.2 Feedback Timing for Reflection

In the traditional studio-based model of design education, students often get feedback both during (via "desk crits") and after work sessions [7, 23]. In fact, as the demand for design education increases, researchers have explored obtaining feedback through online crowds and communities [39, 54, 59, 63, 65, 71, 79, 109]. Others explored the potential for embedding feedback directly into the design tools [38, 64, 80], bringing in the perspective of feedback timing. Schön describes the distinction between reflection-in-action and reflection-on-action due to their differences in timing relative to a design episode—some reflection happens while in the act of performing a task and other times creators reflect back on their actions that already took place (i.e., a posteriori) [15, 41, 93]. He characterizes reflection-in-action as occurring through "back-talk" with the design situation, where the designer makes discoveries first-hand through the situated action and cognition from physically sketching and engaging with the materials while designing [50, 55, 96]. Schön further describes how this can also be driven by external feedback, such as through intermediate design reviews on in-progress work. Our in-action feedback interactions are modeled based on this second type of reflection-in-action. This is in contrast to feedback at a final design critique on a completed assignment (for reflection-on-action), which serves as a model for the design of our on-action feedback interaction [93].

We are interested in systematically exploring how the timing of feedback (operationalized as in-action or on-action) impacts the creative process. Feedback timing has been explored in prior work through examining the impact of having timing control on goal setting [61] or limiting resources on reflection [31]. Most related to our work is Bayerlein investigation on the impact of timeliness on feedback effectiveness, comparing "extremely timely" and "timely" on-action feedback [8]. While students showed no preferences in timeliness, they find that upon comparing manual and automatic feedback, that automatic feedback actually improved students' perceptions of the feedback constructiveness [8]. Note that these are both instances of on-action feedback (just at different time intervals), whereas our research seeks to compare this posthoc feedback with in-action feedback.

2.3 Creativity Support Tools with Embedded Design Feedback

Automated feedback enables providing more timely feedback at less effort than relying on expert feedback providers in non-creative domains [8, 36]. However for creative domains, this is particularly challenging, while also especially important. Thus, researchers have explored a range of methods for embedding feedback into creativity tools. Several tools call awareness to options to consider while designing [44], such as providing examples and alternatives related to the users' current design [24, 33, 58, 64, 67, 80, 84, 97, 98, 100, 105], directly authoring design elements [47, 85], or supporting design understanding through the lens of relevant artistic concepts [34, 35, 64, 80].

Several prior research tools also embed automated feedback in a range of design domains [32, 38, 64]. Fischer et al. [38] first explored embedding a "critic" that could facilitate actions towards well-articulated design goals; this research envisioned the critic as a collaborator, with the goal of providing sufficient justification for design decisions, while maintaining the user's agency. Lee et al. [64]'s GUIComp explored embedding computational feedback within a visual design tool, such as visual complexity scores, attention maps, and visual recommendations. While this type of feedback provides guidance on how to view the design, it leaves interpretations up to the user—e.g. understanding what design choices might cause a low visual complexity score. Building on this work, our research probe aims to provide feedback within a tool, but in a manner that helps novices associate high-level design principles with low-level design choices.

Many of these works share common goals: reducing the gap between working on a design and receiving feedback, or studying forms of in-action feedback. We take a step back to understand if these are the correct goals to strive towards, or if there might be limitations to earlier feedback that need to be carefully considered in such designs. For instance, it is unclear how offering real-time feedback interacts with self-reflective practices, i.e. would the constant availability of feedback introduce distractions? How might we best display feedback "during" the creation of design? Taking inspiration from the literature on feedback and reflection, we implement our research probe by enabling high-quality "computational" feedback on demand through a Wizard-of-oz approach [89, 93].

Our wizard setup enables empirically understanding the values and pitfalls of in-action versus on-action feedback within a design tool.

3 DESIGN PROBE: EXPLORING EMBEDDED FEEDBACK IN A VISUAL DESIGN TOOL

Towards gathering empirical data on the impact of timing on feedback, we created a Wizard-of-Oz design probe [75] that simulates embedding "computational" feedback within a visual design tool. To enable the study of different timing conditions, we have two core interaction considerations: to design the interface such that feedback is available on demand, and to have the timing controllable by an experimenter.

We first describe our design goals for the probe (Section 3.1) and the design of the probe interface (Section 3.2). We then report on a preliminary study to observe how novice designers might interact with the in-action feedback and we discuss how this led to changes to the design probe before our summative evaluation (Section 3.3).

3.1 Feedback Design Goals

Taking inspiration from prior research in feedback theory [8, 89, 95], we outline the key design goals of our probe and describe high-level approaches for achieving them:

*3.1.1 [**DG1**] Ensure High-Quality Feedback.* Prior work studies benefits of good feedback [37, 89], and also harms of low-quality feedback [18, 20, 22, 81, 109]. Especially given the availability of the feedback early in the design process, we want to ensure **high-quality** to avoid hindering the user's design process [86, 89].

> *Our approach.* We opt to engage human experts as "wizards" to provide this feedback and avoid conflating results with algorithmic inaccuracies. To ensure quality and consistency, we create scripted templates for wizards to more effectively author feedback.

*3.1.2 [**DG2**] Situate in Context.* Feedback should be **context-aware** and should present the relevant context to better support the user's understanding [50].

> *Our approach.* We leverage the context of the canvas by associating each piece of feedback with a numbered location annotation to help users better identify the location of a specific issue.

*3.1.3 [**DG3**] Avoid Unnecessary Disruption.* Feedback—even when available **on-demand**—should avoid disrupting the user's ongoing design experience [3, 34, 38].

> *Our approach.* We subtly inform users when feedback is available, but give them control to initiate feedback requests—i.e., only displaying feedback content after users explicitly request it [3].

*3.1.4 [**DG4**] Reduce Barriers to Requesting.* Cognitive or social barriers such as evaluation apprehension and intimidation should be minimized when requesting feedback [57, 60, 74]. Interacting with a (seemingly) **computational** tool can mitigate the social pressures of interacting with humans.

> *Our approach.* By presenting the feedback as computational (rather than expert-provided), we aim to capture preferences for feedback in absence of social pressures. To reduce uncertainties around feedback availability, we provide indication of when it is available to be requested.

*3.1.5 [**DG5**] Allow Creative Flexibility.* Similar to Fischer et al. [38], we want to promote user agency over creative decisions [38, 86, 89]. The tool should allow users' expressivity [17, 95] and encourage using their own creative judgment over the tool's suggestions [43].

> *Our approach.* We visualize distinct options to address or dismiss feedback. By giving the option to explicitly disregard feedback, we aim to reduce any pressure to accept all feedback [26, 41].

3.2 Probe Interface

First, we describe the design of our feedback probe[1] (Section 3.2.1) as informed by our design goals (Section 3.1). We then describe additional experimenter tools (Section 3.2.3) for enabling the wizarded feedback and timing differences.

3.2.1 User Interface: Visual Design Tool. The user interface comprises three major sections (Figure 2). The left panel contains necessary tools that the users need to complete a design, such as "Background" and "Text" elements. The canvas in the middle is where users compose a visual design. Finally, the **"Principled Feedback"** panel (right) contains our core features for supporting feedback-related interactions.

- The **Feedback Request** section (Figure 2a) gives users access to feedback. Selecting the "Request Feedback" button shows the available feedback. The purple notification dot (top-right of the button) informs users that feedback is ready. While our feedback was wizarded by humans, the affordance of the button simulates a fully computational tool.
- The **Principle Tabs** (Figure 2b) act as a static reference (akin to a textbook) providing information about each high-level design principle, including a definition and common issues.
- The **Issue Cards** (Figure 2c) are expandable text areas that present the tool's feedback. The "Resolved" and "Dismiss" buttons (Figure 2h) at the bottom of each card allow discarding feedback by either indicating that an issue has been addressed ("Resolved"), or that they disagree and want to "Dismiss" the issue.

3.2.2 Framework for Templated "Computational" Feedback. To generate high-quality feedback on demand that could be embedded into a computational design tool, we developed a framework based on common principles and issues in visual design. We aimed to structure feedback such that it could be feasibly generated automatically in the future.

Referencing a range of educational design resources [1, 27, 30, 68, 71, 87, 101], our team collectively synthesized an initial set of principles and issues (Appendix: Table 2). We chose to focus on five principles (hierarchy, alignment, balance, unity, and readability), each of which included two or three corresponding issues.

[1] Code and other materials can be found on the project page: https://ejane.me/inactionfeedback.html

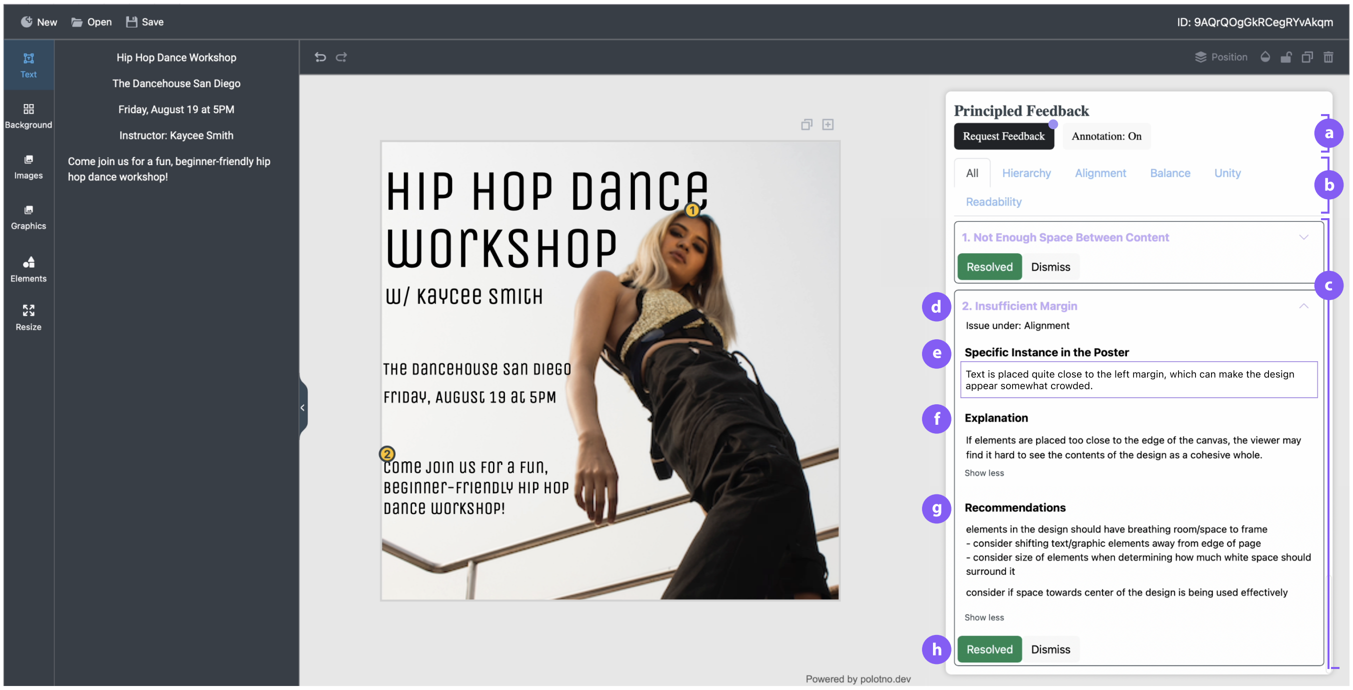


Figure 2: The probe interface is comprised of a tools panel (left), a canvas (middle), and our Principled Feedback panel (right). The feedback panel contains 3 sections: (a) Feedback Request section has a "Request Feedback" button for users to get feedback on their current design (availability shown by the purple dot). To encourage reflection time between feedback requests, this button grays out for a minute after each request (informed by our preliminary study). The "Annotation: On/Off" toggle shows/hides location indicators for each feedback item (the yellow numbered circles on the canvas). The (b) Principle Tabs section provides individual tabs for each principle (Hierarchy, Alignment, etc.) with its definition and associated common issues. Upon requesting feedback, (c) Issue Cards display (initially collapsed for an overview) under the "All" tab. Here, we see two issues. For each issue, the expanded issue card displays (d) the issue name ("Insufficient Margins"), (e) *specific*: a description of the specific instance of the issue in the design ("Text is placed quite close to the left margin, which can make the design appear somewhat crowded.") , (f) *justified*: a higher-level explanation of the issue, (g) *actionable*: recommendations for addressing the issue, and (h) buttons for the user to "Resolve" or "Dismiss" feedback. In this example, the yellow numbers (1) and (2) annotation indicates the locations of the "Not Enough Space Between Content" and "Insufficient Margins" issues, respectively.

To provide feedback on a given issue, we leverage characteristics of effective feedback: specific, actionable, and justified [89]. To turn these characteristics into concrete elements of our feedback, our framework includes content to:

- identify a **specific** issue and ways in which it can be instantiated in the current design,
- **justify** why the issue violates a design principle through explanations, and
- provide **actionable** approaches for addressing the issue through recommendations.

To further facilitate consistency in issue identification and to avoid providing feedback that is overly refinement-focused, we include a "prominence" column defining a minimum requirement for identifying each issue. Figure 2c shows an Issue Card in the design probe authored using the feedback framework. To use the framework, the wizard identifies that the "Insufficient Margin" issue is prominent (margin is less than the width of a character), selects the location of the issue, and adapts the associated specific instance text to the current design. The explanation and recommendations are automatically populated by the framework. The full feedback framework that includes design principles and their corresponding issues, explanations, and recommendations are is provided in supplemental materials.

Our framework—while covering a wide set of principles and issues in visual design—is not an exhaustive list. The tool is designed to allow adding custom principles and issues without significantly influencing the overall interaction.

3.2.3 Experimenter Interfaces. To realize the vision of the design probe without bias, two additional interfaces help experimenters facilitate studies: 1) a **wizard interface** allows a human wizard to construct the feedback while blind to study conditions, and 2) an **interviewer interface** allows the interviewer manage condition-specific logistics.

- In the **wizard interface** (Figure 3), the right side reflects real-time changes on the user's canvas (Figure 3c). Wizards

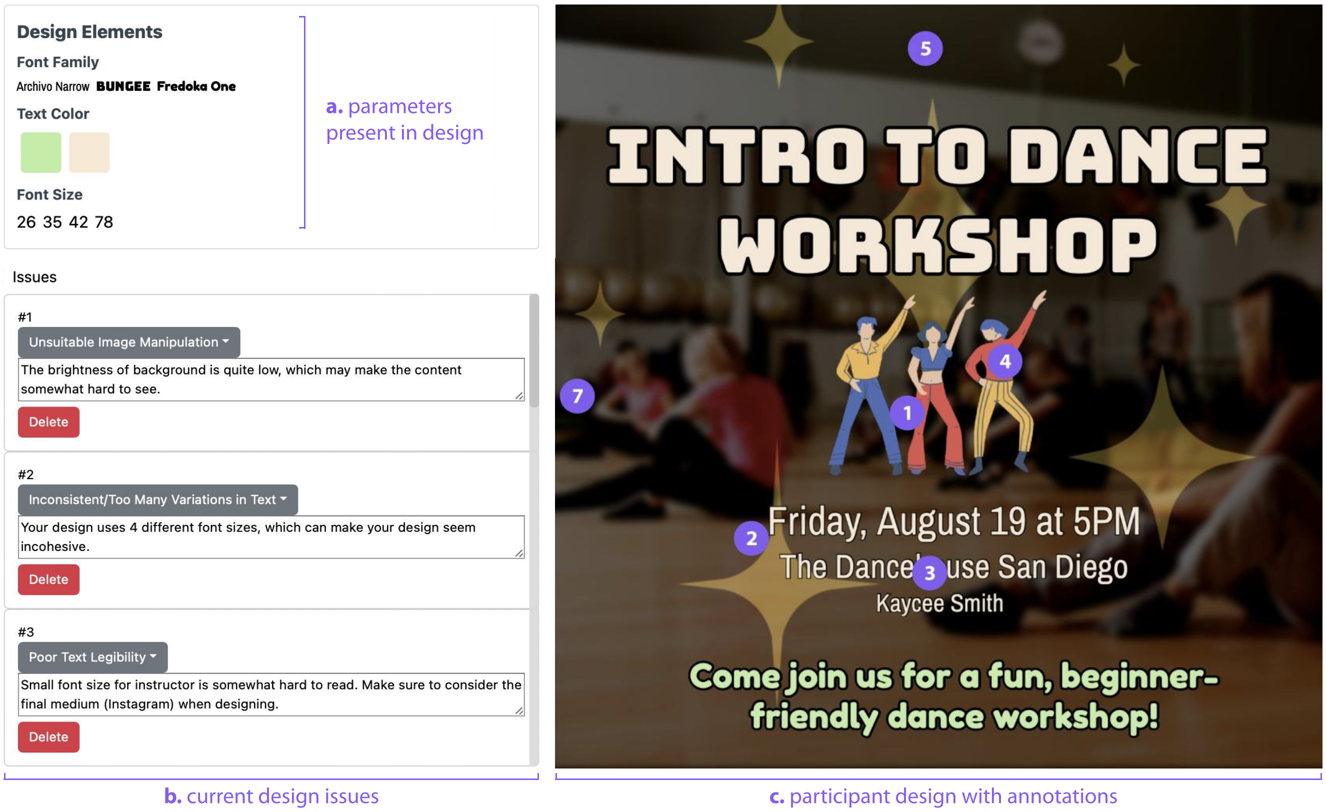


Figure 3: The wizard interface provides support for the wizard to annotate issues within the user's design. The interface shows (a) design parameters present in the current design, (b) an area for authoring new issue information (where we can currently see 4 of the wizard's annotated design issues), and (c) a canvas reflecting the user's current design where the wizard can click to mark the corresponding location of an issue.

can click to annotate a position for their feedback (the location annotation as shown in Figure 2). An empty issue is created in the issue authoring section (Figure 3b). The wizard selects the issue from a dropdown, and customizes the text to describe the specific issue. The top-left section labeled "Design Elements" (Figure 3a) shows text attributes in the user's design to help identify issues pertaining to unity and readability. As the wizard authors feedback, it also appears in the interviewer interface (Figure 4b).

- Our **interviewer interface** (Figure 4) supports managing study logistics. The interface lists the wizard-authored issues, with "Publish" and "Unpublish" buttons, allowing the interviewer to determine when pieces of feedback are made available to the user.

3.2.4 Implementation. Our main interface was built using Polotno SDK v1.0.0-5[1], an open-sourced visual design editor, customized with JavaScript libraries and React components. We trimmed down their default web-based canvas editor to a basic editor with only the necessary tool bars for creating a design (removing extra sections like "Templates"). We built on top of this bare-bones editor to directly embed our "Principled Feedback" panel. To enable synchronicity across the user and experimenter interfaces, we maintain the shared state of the canvas and feedback in a Firebase database[2].

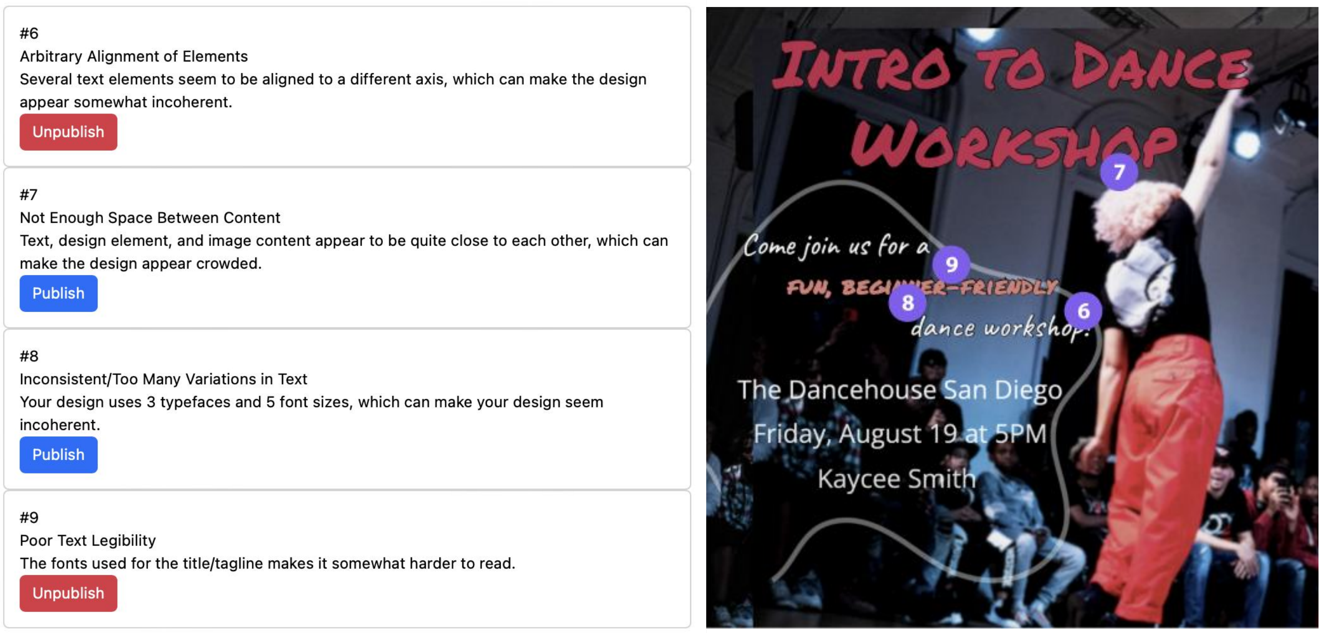
3.3 Preliminary Study: Probe Evaluation and User Reaction to In-Action Feedback

We conducted an initial evaluation to validate our probe design. While our probe design is informed by a theoretical framework (Section 3.1), we wanted to observe the behaviors that arise when feedback is integrated into a design tool. Our focus was on in-action scenarios, as previous research has extensively examined on-action feedback. Our approach enables uncovering potential design considerations that may not have been apparent for our initial design goals.

3.3.1 Method. We ran an initial Wizard-of-Oz study with 11 novice visual designers (5F, 6M; age: 20–32, $\mu = 27.5$) where they used the in-action feedback tool for 20 minutes (P1–P11). Participants were given a design brief to create a square visual design for advertising an upcoming dance workshop. The one-hour sessions

[1] https://polotno.com/

[2] https://firebase.google.com/products/firestore



a. current design issues **b.** participant design with annotations

Figure 4: The interviewer interface provides support for the interviewer to make current issues provided by wizards visible to users. The interface shows (a) current design issues with "Publish"/"Unpublish" options and (b) the annotated user design canvas. At this moment if this user were to request feedback, they would see the two pieces of published feedback: "#6: Arbitrary Alignment of Elements" and "#9: Poor Text Legibility."

were conducted on Zoom and screen-recorded to capture participant behaviors and think aloud. Participants were paid $15 per hour for their time. This early test also helped authors to practice collaboratively managing the task of wizarding the feedback.

3.3.2 Findings. We summarize key observations and participant feedback from the preliminary study.

Participants Appreciated Assistance from In-Action Feedback. Participants appreciated the feedback and addressed them in their designs in meaningful ways (10 of 11). P5 expressed that "the suggestions or hints gave me validation for addressing the issue, encouraging me to experiment". P6 described the tool as "a more experienced collaborator who helped [them] make better decisions, taught [them] new vocabulary, and helped [them] articulate my decisions and issues." They also still felt in control of their creative process due to taking things as no "more than a suggestion" (P4).

Explicit Feedback Options Promote Creative Freedom. Feedback from our participants highlighted how providing feedback control (i.e. when to receive feedback and how to respond) greatly influenced their perception of the tool. Most participants (8 of 11) expressed that they still felt creative freedom and ownership over their designs. Many directly attributed this to being able to "dismiss or resolve and can ask for feedback whenever I want" (P2). Another participant appreciated the proactive requesting behavior, claiming it was "important that I am the one who made the request" (P1).

Feedback Timing and Frequency May Promote Overreliance. Users expressed timing preferences for when they started receiving feedback. Most participants (9 of 11) communicated reasons for not wanting feedback too early: primarily, not being ready (3) and wanting to have a basic draft first (7). Several participants described feeling too focused on addressing feedback: "feedback might interfere with my current thought [...] like someone watching over my shoulder" (P3), "Got stuck focusing on the areas highlighted by the system and forgot about other elements of the design" (P5), and "felt like the design process was gamified and was encouraged to 'win' or resolve all issues" (P7).

In particular, frequent requesting behaviors enabled reliance on the tool's opinions over determining their own creative preference—once feedback became an unlimited resource, participants felt less of a need to self-evaluate [31]. A behavior in participants (5 of 11) that exemplified this reliance was a tendency to make feedback requests in rapid succession towards the end of the task. For example, P10 requested feedback 18 times (of 20 total) within the last 5 minutes of the design task, while engaging in cycles of trial-and-error: testing small design variations and checking, for each, if the tool thought it was a valid solution.

3.3.3 Study and Interface Design Changes. In response to participants' feedback and behaviors, we iterated on both the design probe interface and the study design to more effectively support embedded in-action feedback.

- Given the relative eagerness to address feedback, we softened the language in our feedback templates to **avoid being prescriptive** (**DG5**). Specifically, we refined the language to emphasize that these were merely "suggestions" to help maintain the users' control over the design process.
- Participants appreciated options for requesting, resolving, and dismissing feedback, despite the added interactions. This validated our choice show feedback **only upon request** to avoid being disruptive (**DG3**).
- We adjusted our wizarding behaviors to **wait for an initial skeleton** that fulfills the design task before providing feedback to avoid early distractions (**DG3**). To support this change, we modified our study task instructions to ask participants to first focus on drafting a skeleton design before moving on to refining.
- To further reduce feelings of being overwhelmed by feedback, we also **limited the amount** of feedback presented at a single request to three. Overloading participants with feedback could cause them to ignore feedback content or impact their ability to make sense of the feedback content [40, 108].
- Participants appeared rushed to complete their designs. We therefore **extended the design task time** to see how the tool would be used without the time pressure.
- Perhaps our most important consideration in modifying our design probe was: how to provide **in-action feedback without encouraging overreliance**? Reducing barriers to feedback requests (**DG4**) was inadvertently encouraging regular disruptions (**DG3**) and we wanted to find a better balance. To discourage rapid requesting behavior, we added a 1-minute cool down period for the "Request Feedback" button: during this minute, the button is grayed out, signaling that it is inactive to **encourage down time between requests** during which users can reflect on their design.

4 STUDY: IN-ACTION VERSUS ON-ACTION FEEDBACK

How does the feedback availability impact process, outcomes, and attitudes? While the preliminary study provided initial evidence that novice designers appreciate in-action feedback, we wanted to directly compare this earlier feedback practice to the more common current practice where designers typically only obtain feedback after a design session, or on-action feedback. To do so, we explore how participants perform a design task using our tool when provided either in-action feedback or on-action feedback.

Through this study, we aimed to answer:

> **[RQ1]** How do participants behave differently with access to in-action versus on-action feedback?
> **[RQ2]** What design outcomes changed due to the availability of in-action versus on-action feedback?

We analyze participants' behaviors and interviews, and recruit expert designers to evaluate the quality of participants' designs. We additionally investigate the design implications of in-action feedback for creative learning.

4.1 Study Conditions

All participants worked in the same visual design tool with access to the same feedback panel. In our representations of in-action and on-action feedback (Figure 1), we highlight differences in timing of feedback availability. We describe how we took inspiration from Schön (Section 2.2) to operationalize the two feedback mechanisms within the context of a single design episode:

- **In-action feedback** is feedback available *throughout* the design session on the current action. Participants could access the feedback panel as many times as they wanted after creating an initial skeleton draft. We model this interaction after the intermediate design review described by Schön [93]. Users are able to request mini design reviews as they are in-progress working on a design.
- **On-action feedback** is feedback *after* a completed design on action that already occurred. Participants in this condition were given one opportunity to request feedback after they finished a draft, inspired by the design critique after a completed assignment. Participants were informed at the start of their single feedback and revision opportunity. To further emphasize "completion" of a first design, we asked them to save this first design and send it to us via Zoom.

Conditions were otherwise identical, including number of feedback instance displayed per request (max 3). While this choice may result in variation in the total amount of feedback exposure between conditions, we believe this is true to the two feedback mechanisms.

4.2 Experimenter Roles

Each study had two experimenters: a wizard, who authored feedback, and an interviewer, who managed study conditions.

The **wizard** continuously identified all prominent issues as described by our feedback framework once participants completed a skeleton design (Section 3.2.2). To ensure consistency in feedback quality across conditions, the wizard was blind-to-condition. Wizards all had course-training in visual design and had additionally collaborated in providing feedback in the preliminary studies as training to further encourage consistency.

The **interviewer** walked the participant through the study. They were aware of the study condition and played the administrative role of determine when and what feedback was presented to participants. For on-action participants, the interviewer waited for the participant to indicate that they were done with a draft before publishing feedback. For in-action participants, they could publish feedback as it appeared on their interface. The interviewer throttled the amount of feedback (at most 3), prioritizing prominence and coverage of issues.

4.3 Participants

We recruited 20 participants (15F, 5M; age: 19–60, $\mu = 24.1$) for a between-subjects study with 10 participants in each condition (on-action: OA1–OA10; in-action: IA1–IA10). Participants were recruited through SONA and varied design forums. All participants self-identified as novices with an interest in learning visual design.

on-action	# seen	# unique	# principles	# address	% address
OA1	3	3	3	1	33%
OA2	3	3	3	0	0%
OA3	2	2	2	2	100%
OA4	2	2	2	2	100%
OA5	3	3	3	2	67%
OA6	3	3	2	2	67%
OA7	3	3	3	2	67%
OA8	3	3	3	2	67%
OA9	2	2	2	1	50%
OA10	3	3	2	2	67%
average	2.7	2.7	2.5	1.6	62%

in-action	# seen	# unique	# principles	# address	% address
IA1	7	6	4	6	86%
IA2	6	5	4	5	83%
IA3	6	6	4	6	100%
IA4	3	3	1	3	100%
IA5	6	4	3	6	100%
IA6	7	6	5	5	71%
IA7	7	5	3	4	57%
IA8	12	8	4	10	83%
IA9	6	6	3	5	83%
IA10	8	7	5	7	88%
average	6.8	5.6	3.6	5.7	85%

Table 1: A table describing users' interactions with feedback during the studies. The table enumerates the number of pieces of feedback seen, how many unique issues were, how many different principles were covered. To illustrate how users' responded to the feedback, we also include how many the participant agreed with and tried to address, as well as the percentage of feedback addressed. All columns were statistically different between the in-action and on-action condition studies.

4.4 Procedure

Study sessions lasted around 1.5-2 hours and were conducted over Zoom. Sessions were screen-recorded to capture participant behaviors and think-aloud comments. Participants were paid $30 for their time. All study materials (protocol, surveys, interview questions, etc.) are provided as supplemental materials.

4.4.1 Pre-Test. As a learning measure, we tested participants' ability to identify issues before and after performing the design task. Participants were asked to evaluate a poster design. They were given a reference table of design principles, issues, and definitions (matching the Principle Tabs, Figure 2b) and asked to select each issue they would present as feedback on the design. The provided design had 5 of the most common issues present in designs from our preliminary study.

4.4.2 Interface Walk-Through. Participants were introduced to the tool starting with a walk-through of the static Principle Tabs. Next, they were guided by the interviewer to perform a set of steps on the canvas to familiarize them with the Polotno's canvas editor. Wizards gave pre-written feedback on the resulting design. Once this feedback was provided, the wizard left for the design task, and the interviewer introduced remaining feedback interactions.

4.4.3 Design Task. Participants were given an hour for the design task. They were asked to think out loud and to use the task as a learning opportunity. The task involved designing an Instagram post to advertise a dance workshop. It was designed to be relatively simple, while having enough elements to bring in each design principle. Our task involved multiple text elements of varying importance—introducing hierarchy and unity across content of similar importance. It had one longer tagline and required a main graphic, requiring considerations of readability in terms of text size and placement. Positioning elements relative to each other introduced considerations of alignment and balance.

4.4.4 Post-Test. This was identical to the pre-test but for a new poster design with a different set of common issues.

4.4.5 Survey. Our post-task survey consisted of the Creativity Support Index [17] and usability metrics around whether feedback was useful, efficient, communicative, and satisfying. We additionally included some AI usability metrics such as if feedback was predictable, comprehensible, or controllable, inspired by Oh et al. [85].

4.4.6 Interview. We asked participants to reflect on their experience in an open-ended interview. We asked how they felt about the feedback panel, how the feedback impacted their design process, if it inhibited their creative freedom in any way, etc. In a retrospective, we showed participants all of their wizard-authored feedback. For each, they were asked if and how they addressed the feedback if it was seen (or if they would've addressed it if not seen).

4.5 Analysis Methods

We collected screen-recordings of the Zoom, survey answers, notes from the interviews, and user interaction logs with the Principled Feedback panel.

4.5.1 Log Analysis. Logs recorded interactions with content in the Principled Feedback panel. These were cleaned for feedback requests prior to a skeleton (as no feedback was being generated) and any "double-click" type interactions and analyzed to understand how users interacted with feedback.

4.5.2 Design Quality. To understand the progress of the participants' designs, we annotated a skeleton design per participant—defined as the initial design where all required content is present and the participant's changes have briefly settled. Two external experts with professional visual design experience were recruited to evaluate the participants' designs (in randomized order) at these 3 stages. Experts were asked to identify issues in each design.

4.5.3 Participant Quotes. Authors collaboratively synthesized interview quotes into themes using affinity diagramming [70]. Quotes were sorted exhaustively into distinct clusters by idea.

4.5.4 Iteration Metric. We coded iteration between pairs of designs based on changes across a set of design parameters: background color, layout (text), layout (image), text (font), text (other), and other. Coding was binary per parameter: either changes were made along this dimension or not. Two authors coded these independently (in randomized order) and reviewed together to resolve inconsistencies.

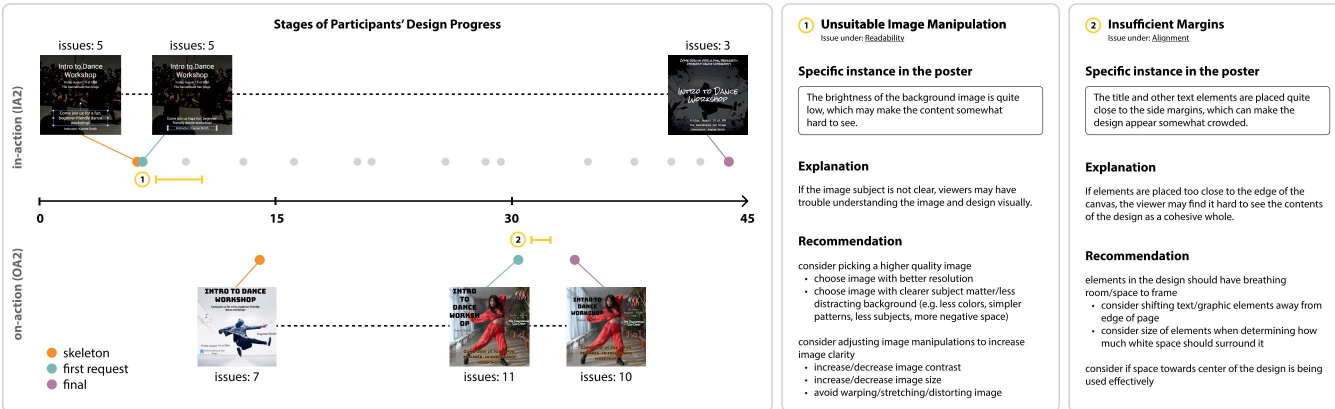


Figure 5: Illustrates the timeline for two participants: one in the in-action condition (top) and one in the on-action condition (bottom). The right two boxes show examples of two feedback panels that participants received and addressed, (1) for the top participant (IA2), and (2) for the bottom participant (OA7). The dots on the design task timeline illustrate when they had a basic skeleton (orange), made their first feedback request (teal), and completed their final design (purple). The additional gray dots for the in-action participant show subsequent feedback requests. For this in-action participant, the skeleton and first request design visually look similar and occur close in time, whereas the on-action participant's first request design is quite similar instead to the final design.

4.6 Results

Participants from both conditions enjoyed using the feedback features of our design probe, agreeing that the tool was usable: useful, satisfying, predictable, comprehensible, etc. Fifteen participants (8 in-action; 7 on-action) mentioned that the tool helped them be more aware or "mindful" (OA10) of design principles: "a lot of my design knowledge came from looking at things and seeing what is good. [The tool] made me aware of which [principles] I was lacking" (IA5). Two participants (1 in-action; 1 on-action) even suggested continued impact on future design projects: "I learned design principles that I will keep in mind in the future" (IA9). IA4 describes gaining a stronger grasp of how to apply principles to designs: "I probably would've had no clue where to start [...] so it was helpful to learn principles and see how it plays in practice with the tool." IA10 described feeling more confident in her own evaluation: "seeing something that I already was thinking about as feedback, felt like reinforcement that I was able to notice these mistakes." In fact, OA2 mentioned the tool being "helpful [for] learning but not for [creating a] better design."

4.6.1 [RQ1] How do participants behave differently with access to in-action versus on-action feedback?

Quantifying On-Action Versus In-Action Feedback Experience. Here we summarize some differences in the participants' exposure to feedback across the conditions both to assist with interpretation of results as well as to characterize differences in behaviors when provided access to in-action versus on-action feedback (Table 1). Figure 5 shows two example timelines from one in-action and one on-action participant.

In the in-action condition, participants made 2–16 feedback requests ($\mu = 9.6$, $\sigma = 4.1$). In fact, several participants (3 of 10) saw every piece of feedback provided (only 1 on-action). Overall there were up to 5 pieces of feedback that weren't seen by participants in both conditions (in-action: $\mu = 1.2$, $\sigma = 1.5$; on-action: $\mu = 3.4$, $\sigma = 2.0$). As expected, in-action participants generally saw more feedback. They saw between 3–12 pieces of feedback ($\mu = 6.8$, $\sigma = 2.3$) versus 2–3 in the on-action condition ($\mu = 2.7$, $\sigma = 0.5$). Of these issues seen by in-action participants, 3–8 were unique ($\mu = 5.6$, $\sigma = 1.4$) and for on-action participants, all were unique. This greater diversity in feedback issues also gave the in-action participants more exposure to different design principles. The feedback spanned 1–5 different principles for in-action participants ($\mu = 3.6$, $\sigma = 1.2$), and 2–3 for on-action participants ($\mu = 2.5$, $\sigma = 0.5$).

In total, wizards provided more feedback (seen and not seen) on designs made in the in-action condition studies [in-action: $\mu = 8.0$, $\sigma = 2.7$; on-action: $\mu = 6.1$, $\sigma = 2.4$; $t(18) = 1.77$, $p < .05$], possibly suggesting that more changes were occurring on the canvas causing more turnaround in identified issues (old issues addressed, new issues created).

In-Action Participants Iterated More Based on First Feedback. We saw evidence that earlier timing of feedback, regardless of quantity, does have greater potential for inspiring iteration. To isolate the impacts of feedback timing from frequency and quantity, we looked at changes in participants' designs following their first feedback request. For on-action participants, we compared their design at the first (and only) feedback request to their final design (Figure 6). For a similar comparison for in-action participants, we used the average time on-action participants spent revising (411sec) as a fixed time interval, comparing their first (of many) request design with the design after this amount of time passed.

To further validate that these intervals are comparable in feedback quantity, we compare feedback counts between on-action and in-action participants. At the first request, in-action participants actually received less feedback than on-action participants [in-action: $\mu = 1.4$, $\sigma = 0.70$; on-action: $\mu = 2.7$, $\sigma = 0.48$; $t(18) = -4.84$, $p < .0001$]. Since in-action participants could continue to request

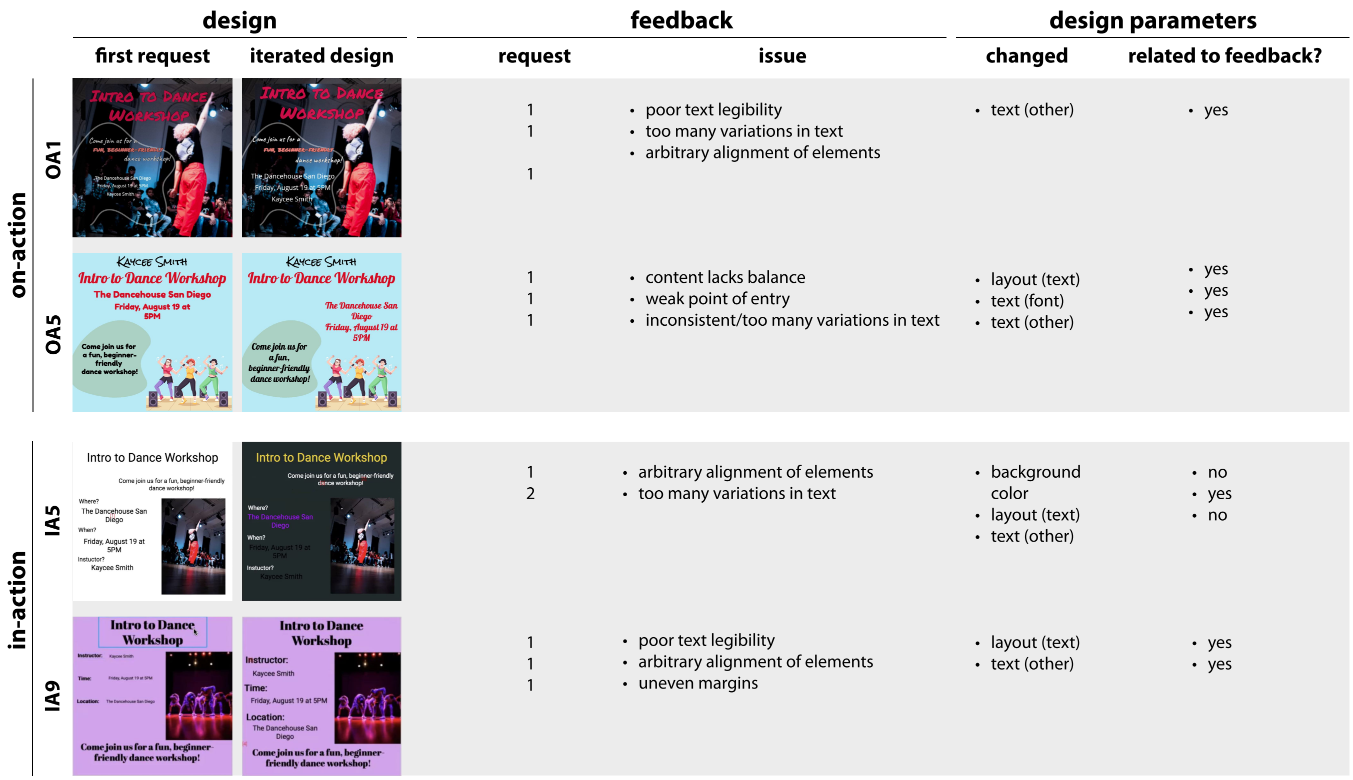


Figure 6: Selected pairs of poster designs from participants in the on-action (top) and in-action (bottom) feedback conditions to showcase iteration following their first feedback request. The first column contains the design at their *first request*, and the third column contains their *iterated designs*. In the on-action examples, these are the final designs. For the in-action participants, these designs were at a fixed time interval (equal to the average time between feedback request and final design for on-action participants, $411sec$) after their first request design. The middle columns list the feedback received between the requests. For in-action participants, some of the *issues* were presented at additional feedback requests, the corresponding *request* count is also provided for content. Finally, the right two columns list the design parameters that were coded as having *changed* between the designs and whether or not these parameter changes were directly *related to feedback* that was received.

feedback during this time, we additionally counted the total number of pieces of feedback seen, which was still less than on-action participants ($\mu = 2.5$, $\sigma = 1.35$).

Representing iteration as the number of design parameters that changed, we saw that in-action participants iterated more than on-action participants during this time following the first feedback request [in-action: $\mu = 3$, $\sigma = 1.15$; on-action: $\mu = 1.8$, $\sigma = 0.79$; $t(18) = 2.71$, $p < .01$]. We further coded each design parameter change based on its relationship to feedback received. We found that a greater percentage of the on-action participants' changes were directly related to feedback, suggesting in-action participants were doing more additional iteration to their design outside of addressing feedback [on-action: $\mu = 0.77$, $\sigma = 0.34$; in-action: $\mu = 0.47$, $\sigma = 0.39$; $t(18) = -1.82$, $p < .05$].

In-Action Participants Requested Feedback Soon After a Skeleton. Participants in the in-action condition requested feedback significantly earlier in their design process [in-action: $\mu = 946sec$, $\sigma = 401$; on-action: $\mu = 1560sec$, $\sigma = 784$; $t(18) = -2.2$, $p = .02$] (Figure 7). In fact, 3 participants requested feedback even before a completed skeleton. Their first feedback requests often occurred as soon as they finished their skeleton design (4 of 10) or when they were feeling stuck (5 of 10).

On-Action Participants Self-Check Before Requesting. On-action participants tended to be more hesitant about their single feedback request. Thus, they were prone to more self-checks for mistakes (3 of 10), making sure that they had met the design task guidelines (3 of 10), and achieved satisfaction for their design (5 of 10) before requesting feedback. In-action participants instead tended to rely on the tool, often making a final request right before finishing as a final check (5 of 10 requested within roughly the last minute of their design time).

In-Action Feedback Can Risk Overreliance and Hinder Creative Freedom. Across conditions, participants mentioned feeling creative and feeling ownership, which was further supported by relatively high CSI scores (in-action: $\mu = 218.6$, $\sigma = 56.4$; on-action: $\mu = 209.6$, $\sigma = 43.9$). Participants (6 in-action; 4 on-action) liked having control over feedback interactions: "the 'Dismiss' option was helpful to say 'No, I disagree with this, therefore I will stick to what I want

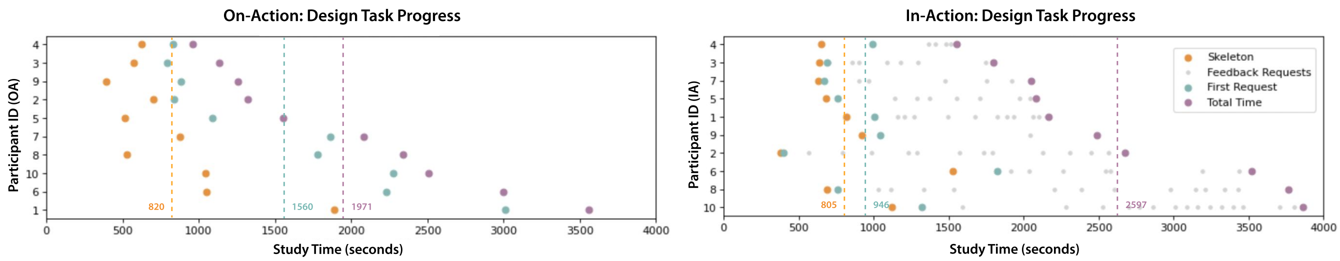


Figure 7: This graph shows the absolute times during the design task when the participant had an initial skeleton (orange), made their first feedback request (teal), made further feedback requests (gray), and completed their final design (purple). Average times are visualized as vertical dotted lines in the corresponding colors. Notice that the first request time is towards the later end of the design session for the on-action condition (left), but happens very soon after the initial skeleton in the in-action condition (right).

to accomplish creatively'" (IA4), or "It's there to help your design improve, you can take it or not" (IA9).

In-action participants mentioned more potential inhibitors, such as iterating solely based on the feedback provided without thinking critically (4 of 10). They felt that they "had less confidence in [their] understanding and ability to point out errors" (IA10), and thus, "trusted [the tool] more than [they] trusted [themselves]" (IA5). Furthermore, two in-action participants mentioned feeling restricted to strictly adhering to design principles, citing examples like feedback suggesting to "center everything, but… in some designs, you don't have to make everything centered" (IA10) or feeling pressured to use fewer fonts or colors. This inclination to "focus too much" (IA3) on feedback should be a consideration for feedback tools aiming to encourage creative freedom.

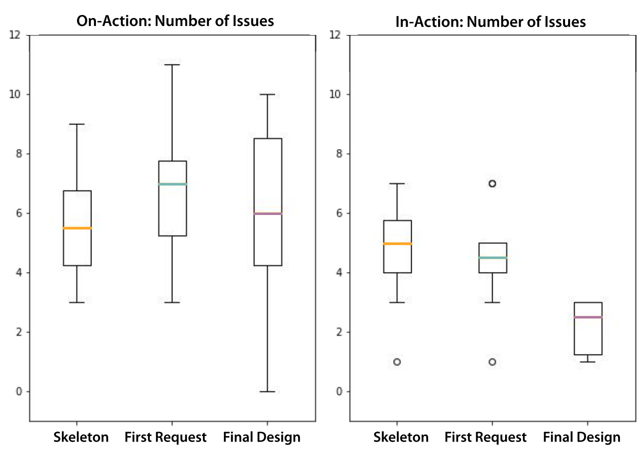


Figure 8: These boxplots show the average number of issues in participants' designs across the different design stages: skeleton (orange), first request (teal), and final design (purple). Notice that in the in-action condition (right), the average number of issues tend to steadily decrease, whereas in the on-action condition (left), participants often created a few issues between the first request and skeleton that weren't necessarily fully addressed after receiving feedback.

4.6.2 [RQ2] What design outcomes changed due to the availability of in-action versus on-action feedback?

In-Action Feedback Encourages Confidence in Ability to Apply Design Principles. Our pre/post-tests showed that in-action participants gained confidence in their ability to identify violations in a design [$U = 21, z = 1.9, p = .03$]. However, the tests did not yield signs of learning: participants did slightly worse in both conditions on the post-test, possibly because content just learned may need time to sink in before it can be used for evaluation [42]. Nonetheless, in-action participants felt more confident in their ability to identify design violations [$U = 21, z = 1.9, p = .03$].

In fact, on-action participants expressed that they would've appreciated in-action feedback throughout their design process. OA5 described being likely to have made different choices if given the in-action feedback: "I kind of wish I'd seen this. Looking back, I probably chose one of the worst options [...] I'd much prefer to fix it." Another participant mentioned, "If I have a chance to request feedback earlier, it would be better. I learned lots about principles after requesting" (OA7).

In-Action Feedback Improved Design Quality. Expert ratings showed improved quality (fewer issues) in both conditions (Figure 8), but the in-action group demonstrated significantly greater improvement [$t(18) = 3.36, p = .002$]. In-action participants showed improvement at both stages: before and after their initial feedback request (before: $\mu = 0.2, \sigma = 1.0$; after: $\mu = 2.3, \sigma = 1.4$). Perhaps unsurprisingly, on-action participants fixed fewer issues ($\mu = 0.8, \sigma = 1.1$) in response to feedback. However, we also observed that they were creating more issues (an increase of 1.2 issues on average, $\sigma = 1.8$) after their skeleton design as they were approaching their initial "completed" draft.

Figure 9 shows issues present across different stages of several participants' design processes. In-action participants directly attributed some of these improvements to the feedback. Many (6 of 10) mentioned the tool helping make them aware of an issue and guiding them towards an improved design: "I think overall [the tool] helped me create a better design. If I hadn't gotten the feedback... I wouldn't have made that change" (IA3). On the other hand, 3 of 10 on-action participants expressed feeling like their design experienced little to no change based on feedback.

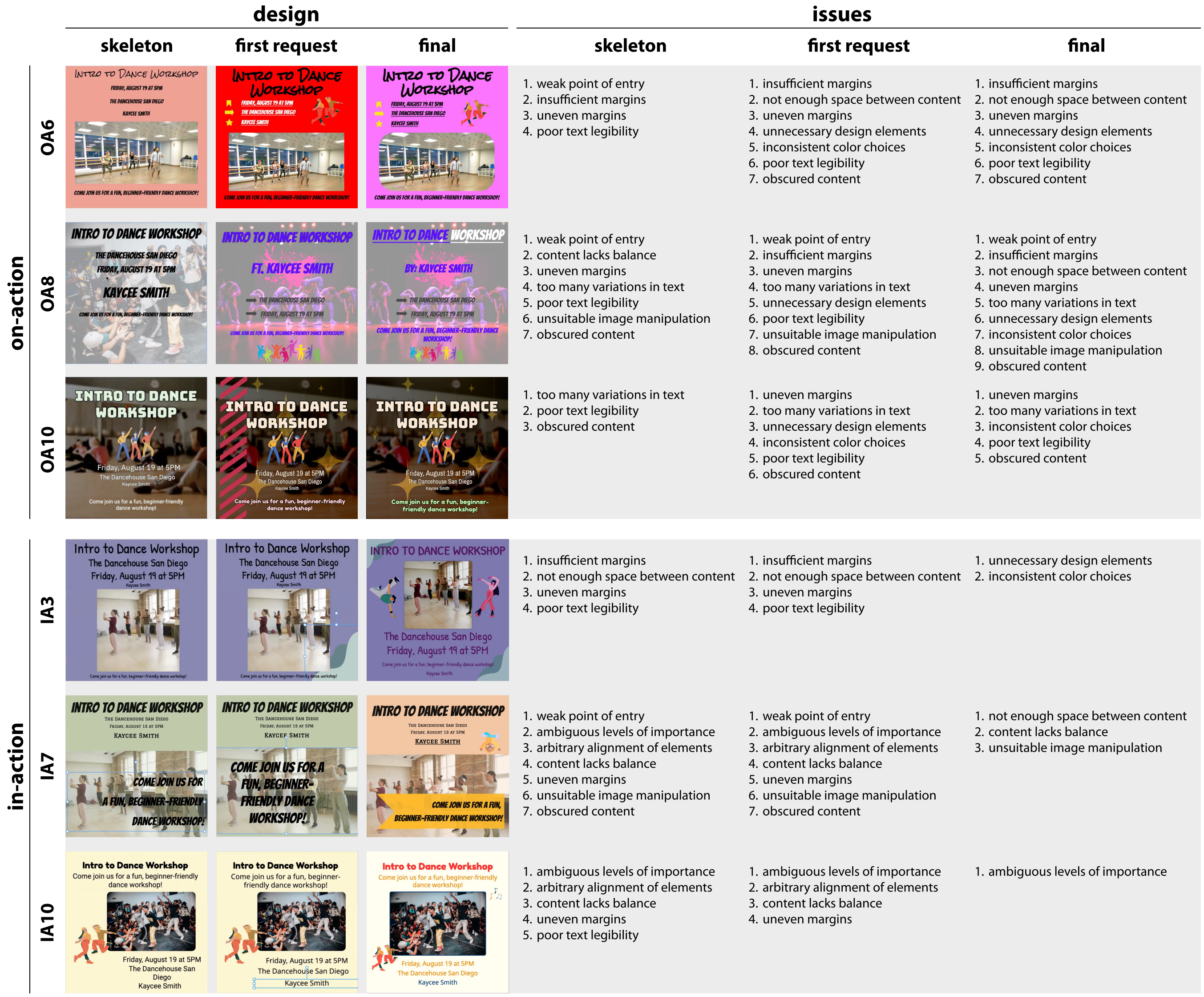


Figure 9: Selected sets of poster designs from participants in the on-action (top) and in-action (bottom) feedback conditions to showcase changes in quality (based on number of issues). The first column contains the *initial skeletons*, the second column contains the design at their *first feedback request*, and the third column contains their *final designs*. The right three columns list the corresponding issues per design, respectively. Notice that first request and final designs are more similar for on-action participants and vice versa for in-action participants.

Design Fixation Limits On-Action Impact on Design Quality. In addition to seeing less feedback, on-action participants were less likely than their in-action counterparts to address the feedback [on-action: $\mu = 0.62$, $\sigma = 0.29$; in-action: $\mu = 0.85$, $\sigma = 0.14$; $t(18) = -2.29$, $p = .02$] (Table 1). This may have contributed to the differences in quality improvement. Of issues participants did receive feedback on, a greater portion remained in on-action participants' final designs [on-action: $\mu = 0.68$, $\sigma = 0.34$; in-action: $\mu = 0.25$, $\sigma = 0.23$; $t(18) = -3.30$, $p = .002$].

In the on-action condition, participants tended to demonstrate reluctance to accept feedback, suggesting fixation: "It would be better if I received [feedback] earlier. If it's [given] earlier, I would think about it more and not be that defensive" (OA7). While addressing a piece of feedback, she even mentioned only being willing to make refinements, "I will try a little bit, but not much."

5 DISCUSSION & FUTURE WORK

In summary, we conducted a mixed-methods between-subjects study with 20 participants who used our design probe to create a visual design while having access to either in-action or on-action feedback from a hidden Wizard who was an expert in design. From analyzing participants' behavior logs, expert evaluations of the participants' designs, and interview transcripts, we discovered some benefits and risks in both in-action and on-action feedback. We found evidence that early feedback (in-action) was more likely to be considered when iterating on the designs, resulting in more improvement in design quality. However, the constant access to this feedback can allow participants to rely on feedback in ways that otherwise would've required more self-reflection. On the other hand, on-action participants were more likely to exhibit fixation and stick with their designs, resulting in fewer and more minor changes after feedback [52]. We discuss factors that may contribute to participants' willingness to collaborate with this "computational" tool, as well as the potential of in-action versus on-action feedback for encouraging learning.

5.1 Implications of Computational Feedback as an Intelligent Co-Creator

As creativity support tools become more intelligent, they have the potential embed knowledge that could provide access to access to basic creative education. To achieve this vision, future tools need to account for aspects of power and trust of the tool and how they might influence user agency and learning.

5.1.1 Power, Agency, and Cognitive Models of Human-AI Co-Creation. Our results further contribute to the discussion of power and the role of AI in creativity support tools [53, 66]. In particular, it highlights the need to strike a balance between AI and human (user) agency, a relationship that changes over time, across tasks, type of user, etc. [73, 90]. Across all participants in our study, we saw evidence of the three cognitive models—offloading, , co-creative partner, and casual creator—observed by Gero and Chilton [47] in the context of collaborative writing with an AI. We saw evidence of **cognitive offloading** in the way that many in-action participation would continually check for new feedback after small change to their design. Some seem to rely on insight from the tool, rather than thinking deeply about their situation. While participants were initially only intending to offload the evaluation aspects of designing to the tool, the extent to which feedback influenced the design process then often meant the tool was more like a **co-creative partner**. Participants who realized this often felt the tool was playing an outsized role and inhibiting some of their creative ownership. Finally, the appreciation participants expressed in the feedback's ability to help call awareness to overlooked issues aligns with the feedback being viewed as a **casual creator** or as one participation described, the tool was like an "experienced collaborator" (P6), able to to help facilitate further learning, exploration, and discovery.

5.1.2 Trusting Automation Can Lead to Overreliance. Overly trusting an AI is a known concern across different types of human-AI collaborative tools [9, 10, 13, 99]. Especially in creative collaborations with generative AI tools, it is important to balance reliance and ownership [12, 47, 85]. In our case, the tool does not generate new content within the design, but is influencing the designers' creative process by guiding their attention through feedback. This indirect impact on creative ownership might further result in reduced awareness of this tendency to overly rely and trust the tool.

5.1.3 Ability to Disregard a Machine's "Opinion". While our tool is not actively "creating," it is acting as a collaborator through its feedback. Prior work has talked about the importance of the user taking the lead in human-AI collaborative creative activities [12, 47, 85]. We also observed participants feeling like the tool's feedback is easier to ignore than a human's, often because they doubted a machine's ability to understand their creative intentions. As in prior work, we saw evidence of correlation between this reduced credibility and dismissing the tool's feedback [12]. While this reduced credibility meant less potential impact of the feedback, it also allowed for more creative flexibility. Note that our observations of overreliance occurred even within this "computational" context, where social factors such as authority are alleviated.

5.1.4 Tools Can Provide Users More Explicit Control. We are interested in additional ways in which leveraging the "computational" nature of a tool can go beyond the normal social interactions with a human expert [51] (beyond reducing social apprehension). We observed that seemingly small design choices to include explicit affordances for supporting user control were empowering [45, 49, 83]. For instance, the explicit presentation of having a feedback request button, and having options for "resolve" and "dismiss" made participants feel like they had more flexibility and ownership over their design choices.

5.2 Implications of In-Action Feedback on Supporting Learning

As noted, in-action feedback potentially emphasizes performance over learning. However, the overall improvement suggests that the in-action participants were able to use the feedback to address issues, which is a step in the direction of learning the principles of visual design [42].

5.2.1 Comparisons to Existing Feedback Interventions. We explore how our performance results may relate to research on a range of feedback mechanisms. Cutumisu and Schwartz [21]'s studies of the impact of critical feedback showed a positive relationship between creative performance and quantity of critical feedback, but only when participants were given the choice between critical (positive) and confirmatory (negative) feedback. While our participants were not given a choice of what kind of feedback they received, the ability to determine when to request feedback seems to have replicated this association between performance and feedback quantity, as in-action participants saw a larger amount of feedback. Krishna Kumaran et al. [61] shows the benefits of planning proximal feedback goals on quality improvement. The availability of our feedback likely similarly encouraged this earlier seeking of feedback, allowing us to see some of the similar benefits of encouraging participants to share early drafts to receive feedback. Finally, in-action participants' increased exposure to feedback could have instigated the learning benefits of rapid iteration, as observed in Dow et al. [31].

5.2.2 Adaptive Feedback Reduces Need for Static References. We provided participants with two learning references, the adaptive feedback and the static Principle Tabs. Engagement with Principle

Tabs can be interpreted with more self-initiated learning of the related design principles, as participants had to make the choice to actively seek out additional knowledge about the design principles on their own. However, across conditions we observed very little interaction with the Principle Tabs, despite some participants expressing interest in the preliminary study. Arguably, adaptive feedback helps participants more directly transfer the design principles, and thus obviated the need for participants to engage with general representations of knowledge. [43, 80]. Of the few Principle Tabs interactions, some behaviors we observed included using them as "a check at the beginning and end" (P1), aligning with known learning behaviors of using checklists as scaffolding [11, 88]. Future work might explore how to balance usefulness of feedback for assisting in knowledge transfer with potentially ambiguity or delay to encourage more self-initiated learning [2].

5.2.3 Terminology Can Influence Feedback Efficacy. Several participants mentioned the language and terminology in the feedback impacting their perceptions of and ability to use the tool. Several participants were very positive and described how learning new terms empowered them to better communicate and support their design intentions in "concrete terms" (OA10). On the other hand, a few other participants mentioned not being able to fully understand feedback and the feedback not being "beginner friendly" (IA10) due to not understanding the words used. Future work might explore how a feedback tool could promote more learning of the relevant terms [4, 72], and facilitate the breakdown when users do not understand the particular terms used [14, 82, 102].

5.3 Study Design Limitations

Our research seeks to understand how these quicker computationally-enabled interactions—such as on-demand feedback—might influence users' creative processes. However, there are many limitations to what we learned due to our study design including: the limited number of participants, the focus on novices, the lab setting, the design of the study conditions, the use of the Wizard-of-Oz methodology for feedback, etc.

5.3.1 Evaluating More Realistic Study Scenarios. We are interested in further understanding the potential of embedded feedback when used for more realistic design scenarios. In particular, we'd be interested in studying its use across a longer period of time. While participants mentioned feeling like they were learning, our pre/post-test yielded no quantitative results. This is likely because the task of evaluation using newly learned concepts, can be difficult for a novice (at the top of Bloom's taxonomy pyramid) [42]. Given a longer period of time, users would be able to learn across multiple design sessions iterating on a one or more designs with personal significance. We wonder if such scenarios might show more learning and less overreliance.

5.3.2 Treating Expertise as a Spectrum. Varying levels of expertise also goes hand-in-hand with these different study scenarios, as more experienced designers might have particular assignments in mind with different learning versus performance outcomes [23, 40, 78]. Future studies expanding the participant pool to include designers with intermediate or advanced expertise could provide additional insights into how feedback timing affects designers at different skill levels, and study how preferences vary across time, process, and type of work.

5.3.3 Comparing Alternatives for On-Action Feedback. In our study, we hoped to simulate a comparison between in- versus on- action feedback while keeping the interactions as similar as possible to avoid introducing external factors. However in doing so, we do minimize the timing differences between conditions, as both are constrained to the time-frame of a single design session [15, 41, 93]. The lack of delay for the on-action condition meant less time to fixate on a design, as well as less intermediate time to reflect on their prior action [93, 94, 107]. Finally, we do not expect that in-action and on-action feedback operate in such a separate manner, but rather we envision them being used together. For instance, in-action feedback might be an additional source of support in the absence of on-action feedback sources. It would be interesting to explore a wider range of on-action, as well as joint (in-action with on-action) feedback scenarios.

5.3.4 Enabling Experimentation on Other Dimensions of Feedback. In designing our "computational" feedback tool, we considered potential extensions to enable a wide range of studies around feedback interactions. The probe comprises components of the full study platform: user-facing design feedback tool, feedback provider (wizard) interface, and an administrative interface for managing study conditions. We were careful to design a template framework that could be reproduced by other wizards, and also easily customized or extended to generate other forms of high-quality feedback. Our experimental design probe has the potential to support further empirical explorations around embedded feedback such as source, presentation style, content, qualitative versus quantitative, personalized versus generic, etc.

6 CONCLUSION

With recent advancements in intelligent tools, the HCI community has sought to understand and design effective interactions for human-AI collaborations. It is becoming increasingly realistic to develop tools that can provide immediate feedback without the need of a human expert. While early (in-action) feedback has a wide range of potential benefits, it is also especially nuanced due to it's potential to impact the designer's creative process from the beginning. We compare in-action feedback to the more traditionally available on-action feedback to understand the tradeoffs in timing on users' creative processes. To explore this, we introduced research probe that embeds feedback within a visual design tool (using a Wizard-of-Oz strategy). Our goal was to understand: within a design environment that can provide on-demand feedback, when do novices want to request it, and how does the feedback impact their design process? Through a between-subjects study with 20 participants randomly assigned to either have in-action or on-action feedback, we discover that novices do tend to find in-action feedback helpful, but that there is a risk that the feedback distracts them from self-initiated reflection. In the context of these potential benefits and risks, we encourage system designers to consider tradeoffs in quality, creative agency, and learning in designing tools with embedded creative feedback.

ACKNOWLEDGMENTS

We would like to thank Teguh Hofstee for helpful discussions around the project idea as well as coding and writing assistance, Sharvari Thippanna and Scott Klemmer for initial explorations on the project, and our participants and reviewers for their valuable feedback. Our research is supported in part by UCSD CSE's Postdoctoral Fellowship and NSF grant #2009003.

REFERENCES

[1] 2022. CorelDRAW Graphics Suite. https://www.coreldraw.com/en/tips/graphic-design-principles/index.html
[2] Susan A Ambrose, Michael W Bridges, Michele DiPietro, Marsha C Lovett, and Marie K Norman. 2010. *How learning works: Seven research-based principles for smart teaching.* John Wiley & Sons.
[3] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N Bennett, Kori Inkpen, et al. 2019. Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–13. https://doi.org/10.1145/3290605.3300233
[4] Lorin W Anderson and David R Krathwohl. 2001. *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives.* Longman.
[5] Frederik Anseel, Filip Lievens, and Eveline Schollaert. 2009. Reflection as a strategy to enhance task performance after feedback. *Organizational Behavior and Human Decision Processes* 110, 1 (2009), 23–35. https://doi.org/10.1016/j.obhdp.2009.05.003
[6] Hal R Arkes and Catherine Blumer. 1985. The psychology of sunk cost. *Organizational behavior and human decision processes* 35, 1 (1985), 124–140.
[7] Terry Barrett. 1988. A comparison of the goals of studio professors conducting critiques and art education goals for teaching criticism. *Studies in art education* 30, 1 (1988), 22–27.
[8] Leopold Bayerlein. 2014. Students' feedback preferences: how do students react to timely and automatically generated assessment feedback? *Assessment & Evaluation in Higher Education* 39, 8 (2014), 916–931. https://doi.org/10.1080/02602938.2013.870531
[9] Natalie C Benda, Laurie L Novak, Carrie Reale, and Jessica S Ancker. 2022. Trust in AI: why we should be designing for APPROPRIATE reliance. *Journal of the American Medical Informatics Association* 29, 1 (2022), 207–212.
[10] Michaela Benk, Suzanne Tolmeijer, Florian von Wangenheim, and Andrea Ferrario. 2022. The Value of Measuring Trust in AI-A Socio-Technical System Perspective. *arXiv preprint arXiv:2204.13480* (2022).
[11] Aditya Bharadwaj, Pao Siangliulue, Adam Marcus, and Kurt Luther. 2019. Critter: Augmenting creative work with dynamic checklists, automated quality assurance, and contextual reviewer feedback. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12. https://doi.org/10.1145/3290605.3300769
[12] Oloff C Biermann, Ning F Ma, and Dongwook Yoon. 2022. From tool to companion: Storywriters want AI writers to respect their personal values and writing strategies. In *Designing Interactive Systems Conference*. 1209–1227. https://doi.org/10.1145/3532106.3533506
[13] Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z Gajos. 2021. To trust or to think: cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–21. https://doi.org/10.1145/3449287
[14] David Carless. 2006. Differing perceptions in the feedback process. *Studies in higher education* 31, 2 (2006), 219–233.
[15] Alberto AP Cattaneo and Elisa Motta. 2021. "I reflect, therefore I am… a good professional". On the relationship between reflection-on-action, reflection-in-action and professional performance in vocational education. *Vocations and Learning* 14, 2 (2021), 185–204. https://doi.org/10.1007/s12186-020-09259-9
[16] Ruijia Cheng, Ziwen Zeng, Maysnow Liu, and Steven Dow. 2020. Critique Me: Exploring How Creators Publicly Request Feedback in an Online Critique Community. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–24. https://doi.org/10.1145/3415232
[17] Erin Cherry and Celine Latulipe. 2014. Quantifying the creativity support of digital tools through the creativity support index. *ACM Transactions on Computer-Human Interaction (TOCHI)* 21, 4 (2014), 1–25. https://doi.org/10.1145/2617588
[18] Kwangsu Cho, Christian D Schunn, and Davida Charney. 2006. Commenting on writing: Typology and perceived helpfulness of comments from novice peer reviewers and subject matter experts. *Written communication* 23, 3 (2006), 260–294.
[19] John Joon Young Chung and Eytan Adar. 2023. Artinter: AI-powered Boundary Objects for Commissioning Visual Arts *(DIS '23)*. Association for Computing Machinery, New York, NY, USA, 1997–2018. https://doi.org/10.1145/3563657.3595961
[20] Maria Cutumisu and Daniel L. Schwartz. 2018. The impact of critical feedback choice on students' revision, performance, learning, and memory. *Computers in Human Behavior* 78 (2018), 351–367. https://doi.org/10.1016/j.chb.2017.06.029
[21] Maria Cutumisu and Daniel L Schwartz. 2018. The impact of critical feedback choice on students' revision, performance, learning, and memory. *Computers in Human Behavior* 78 (2018), 351–367. https://doi.org/10.1016/j.chb.2017.06.029
[22] Deanna Dannels, Amy Housley Gaffney, and Kelly Norris Martin. 2008. Beyond Content, Deeper than Delivery: What Critique Feedback Reveals about Communication Expectations in Design Education. *International Journal for the Scholarship of teaching and Learning* 2, 2 (2008), n2. https://doi.org/10.20429/ijsotl.2008.020212
[23] Deanna P Dannels and Kelly Norris Martin. 2008. Critiquing critiques: A genre analysis of feedback across novice to expert design studios. *Journal of Business and Technical Communication* 22, 2 (2008), 135–159. https://doi.org/10.1177/1050651907311923
[24] Niraj Ramesh Dayama, Simo Santala, Lukas Brückner, Kashyap Todi, Jingzhou Du, and Antti Oulasvirta. 2021. Interactive layout transfer. In *26th International Conference on Intelligent User Interfaces*. 70–80. https://doi.org/10.1145/3397481.3450652
[25] Elly De Bruijn and Yvonne Leeman. 2011. Authentic and self-directed learning in vocational education: Challenges to vocational educators. *Teaching and Teacher Education* 27, 4 (2011), 694–702. https://doi.org/10.1016/j.tate.2010.11.007
[26] John Dewey. 2022. *How we think.* DigiCat.
[27] Donis A Dondis. 1974. *A primer of visual literacy.* Mit Press.
[28] Steven Dow, Julie Fortuna, Dan Schwartz, Beth Altringer, Daniel Schwartz, and Scott Klemmer. 2011. Prototyping dynamics: sharing multiple designs improves exploration, group rapport, and results. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 2807–2816. https://doi.org/10.1145/1978942.1979359
[29] Steven Dow, Elizabeth Gerber, and Audris Wong. 2013. A pilot study of using crowds in the classroom. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 227–236. https://doi.org/10.1145/2470654.2470686
[30] Steven P Dow, Alana Glassco, Jonathan Kass, Melissa Schwarz, Daniel L Schwartz, and Scott R Klemmer. 2010. Parallel prototyping leads to better design results, more divergence, and increased self-efficacy. *ACM Transactions on Computer-Human Interaction (TOCHI)* 17, 4 (2010), 1–24. https://doi.org/10.1145/1879831.1879836
[31] Steven P Dow, Kate Heddleston, and Scott R Klemmer. 2009. The efficacy of prototyping under time constraints. In *Proceedings of the seventh ACM conference on Creativity and cognition*. 165–174. https://doi.org/10.1145/1640233.1640260
[32] Peitong Duan, Jeremy Warner, Yang Li, and Bjoern Hartmann. 2024. Generating Automatic Feedback on UI Mockups with Large Language Models. *arXiv preprint arXiv:2403.13139* (2024). https://doi.org/arXiv.2403.13139
[33] Jane L. E, Ohad Fried, and Maneesh Agrawala. 2019. Optimizing Portrait Lighting at Capture-Time Using a 360 Camera as a Light Probe. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. https://doi.org/10.1145/3332165.3347893
[34] Jane L. E, Ohad Fried, Jingwan Lu, Jianming Zhang, Radomír Mech, Jose Echevarria, Pat Hanrahan, James A Landay, et al. 2020. Adaptive photographic composition guidance. In *CHI*. https://doi.org/10.1145/3313831.3376635
[35] Jane L E, Kevin Y. Zhai, Jose Echevarria, Ohad Fried, Pat Hanrahan, and James A. Landay. 2021. Dynamic Guidance for Decluttering Photographic Compositions. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '21)*. Association for Computing Machinery, New York, NY, USA, 359–371. https://doi.org/10.1145/3472749.3474755
[36] Sean B Eom, H Joseph Wen, and Nicholas Ashill. 2006. The determinants of students' perceived learning outcomes and satisfaction in university online education: An empirical investigation. *Decision Sciences Journal of Innovative Education* 4, 2 (2006), 215–235.
[37] K Anders Ericsson, Ralf T Krampe, and Clemens Tesch-Römer. 1993. The role of deliberate practice in the acquisition of expert performance. *Psychological review* 100, 3 (1993), 363.
[38] Gerhard Fischer, Kumiyo Nakakoji, Jonathan Ostwald, Gerry Stahl, and Tamara Sumner. 1993. Embedding critics in design environments. *The knowledge engineering review* 8, 4 (1993), 285–307. https://doi.org/10.1017/S026988890000031X
[39] Eureka Foong, Steven P Dow, Brian P Bailey, and Elizabeth M Gerber. 2017. Online feedback exchange: A framework for understanding the socio-psychological factors. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 4454–4467. https://doi.org/10.1145/3025453.3025791
[40] Eureka Foong, Darren Gergle, and Elizabeth M Gerber. 2017. Novice and expert sensemaking of crowdsourced design feedback. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–18.
[41] Corey Ford and Nick Bryan-Kinns. 2023. Towards a Reflection in Creative Experience Questionnaire. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–16. https://doi.org/10.1145/3544548.3581077

[42] Mary Forehand. 2010. Bloom's taxonomy. *Emerging perspectives on learning, teaching, and technology* 41, 4 (2010), 47–56.
[43] Krzysztof Z Gajos and Lena Mamykina. 2022. Do People Engage Cognitively with AI? Impact of AI Assistance on Incidental Learning. In *27th International Conference on Intelligent User Interfaces*. 794–806. https://doi.org/10.1145/3490099.3511138
[44] W Timothy Gallwey. 2014. *The inner game of tennis: The classic guide to the mental side of peak performance.* Macmillan.
[45] William W Gaver. 1996. Situating action II: Affordances for interaction: The social is material for design. *Ecological psychology* 8, 2 (1996), 111–129.
[46] Elizabeth Gerber and Maureen Carroll. 2012. The psychological experience of prototyping. *Design studies* 33, 1 (2012), 64–84.
[47] Katy Ilonka Gero and Lydia B Chilton. 2019. Metaphoria: An algorithmic companion for metaphor creation. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–12. https://doi.org/10.1145/3290605.3300526
[48] Katy Ilonka Gero, Vivian Liu, and Lydia Chilton. 2022. Sparks: Inspiration for science writing using language models. In *Designing interactive systems conference*. 1002–1019. https://doi.org/10.1145/3532106.3533533
[49] Rex Hartson. 2003. Cognitive, physical, sensory, and functional affordances in interaction design. *Behaviour & information technology* 22, 5 (2003), 315–338.
[50] James Hollan, Edwin Hutchins, and David Kirsh. 2000. Distributed cognition: toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction (TOCHI)* 7, 2 (2000), 174–196. https://doi.org/10.1145/3313831.3376327
[51] Jim Hollan and Scott Stornetta. 1992. Beyond being there. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 119–125. https://doi.org/10.1145/142750.142769
[52] David G Jansson and Steven M Smith. 1991. Design fixation. *Design studies* 12, 1 (1991), 3–11. https://doi.org/10.1016/0142-694X(91)90003-F
[53] Pratyusha Kalluri. 2020. Don't ask if artificial intelligence is good or fair, ask how it shifts power. *Nature* 583, 7815 (2020), 169–169. https://doi.org/10.1038/d41586-020-02003-
[54] Joy Kim, Maneesh Agrawala, and Michael S Bernstein. 2017. Mosaic: designing online creative communities for sharing works-in-progress. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 246–258. https://doi.org/10.1145/2998181.2998195
[55] David Kirsh and Paul Maglio. 1994. On distinguishing epistemic from pragmatic action. *Cognitive science* 18, 4 (1994), 513–549.
[56] David A Kolb. 1984. Experience as the source of learning and development. *Upper Sadle River: Prentice Hall* (1984).
[57] Yasmine Kotturi and McKayla Kingston. 2019. Why do Designers in the "Wild" Wait to Seek Feedback until Later in their Design Process? In *Proceedings of the 2019 on Creativity and Cognition*. 541–546. https://doi.org/10.1145/3325480.3326580
[58] Yuki Koyama and Masataka Goto. 2022. BO as Assistant: Using Bayesian Optimization for Asynchronously Generating Design Suggestions. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 1–14. https://doi.org/10.1145/3526113.3545664
[59] Markus Krause, Tom Garncarz, JiaoJiao Song, Elizabeth M Gerber, Brian P Bailey, and Steven P Dow. 2017. Critique style guide: Improving crowdsourced design feedback with a natural language model. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 4627–4639. https://doi.org/10.1145/3025453.3025883
[60] Sneha R Krishna Kumaran, Wenxuan Wendy Shi, and Brian P Bailey. 2021. Am I Ready to Get Feedback? A Taxonomy of Factors Creators Consider Before Seeking Feedback on In-Progress Creative Work. In *Creativity and Cognition*. 1–10. https://doi.org/10.1145/3450741.3465255
[61] Sneha R. Krishna Kumaran, Yue Yin, and Brian P. Bailey. 2021. Plan Early, Revise More: Effects of Goal Setting and Perceived Role of the Feedback Provider on Feedback Seeking Behavior. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 24 (apr 2021), 22 pages. https://doi.org/10.1145/3449098
[62] Chinmay Kulkarni, Steven P Dow, and Scott R Klemmer. 2014. Early and repeated exposure to examples improves creative work. In *Design thinking research*. Springer, 49–62.
[63] Chinmay E Kulkarni, Michael S Bernstein, and Scott R Klemmer. 2015. PeerStudio: rapid peer feedback emphasizes revision and improves performance. In *Proceedings of the second (2015) ACM conference on learning@ scale*. 75–84. https://doi.org/10.1145/2724660.2724670
[64] Chunggi Lee, Sanghoon Kim, Dongyun Han, Hongjun Yang, Young-Woo Park, Bum Chul Kwon, and Sungahn Ko. 2020. GUIComp: A GUI design assistant with real-time, multi-faceted feedback. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–13. https://doi.org/10.1145/3313831.3376327
[65] Fritz Lekschas, Spyridon Ampanavos, Pao Siangliulue, Hanspeter Pfister, and Krzysztof Z Gajos. 2021. Ask Me or Tell Me? Enhancing the Effectiveness of Crowdsourced Design Feedback. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–12. https://doi.org/10.1145/3411764.3445507
[66] Jingyi Li, Eric Rawn, Jacob Ritchie, Jasper Tran O'Leary, and Sean Follmer. 2023. Beyond the Artifact: Power as a Lens for Creativity Support Tools. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–15. https://doi.org/10.1145/3586183.3606831
[67] JiayiZhou Li, Junxiu Tang, Tan Tang, Haotian Li, Weiwei Cui, Yingcaui Wu, et al. 2024. Understanding Nonlinear Collaboration between Human and AI Agents: A Co-design Framework for Creative Design. *arXiv preprint arXiv:2401.07312* (2024). https://doi.org/10.48550/arXiv.2401.07312
[68] William Lidwell, Kritina Holden, and Jill Butler. 2010. *Universal principles of design, revised and updated: 125 ways to enhance usability, influence perception, increase appeal, make better design decisions, and teach through design.* Rockport Pub.
[69] J John Loughran. 2002. *Developing reflective practice: Learning about teaching and learning through modelling*. Routledge. https://doi.org/10.4324/9780203453995
[70] Andrés Lucero. 2015. Using affinity diagrams to evaluate interactive prototypes. In *Human-Computer Interaction–INTERACT 2015: 15th IFIP TC 13 International Conference, Bamberg, Germany, September 14-18, 2015, Proceedings, Part II 15*. Springer, 231–248. https://doi.org/10.1007/978-3-319-22668-2_19
[71] Kurt Luther, Jari-Lee Tolentino, Wei Wu, Amy Pavel, Brian P Bailey, Maneesh Agrawala, Björn Hartmann, and Steven P Dow. 2015. Structuring, aggregating, and evaluating crowdsourced design critique. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*. 473–485. https://doi.org/10.1145/2675133.2675283
[72] Stephen MacNeil, Zijian Ding, Kexin Quan, Thomas j Parashos, Yajie Sun, and Steven P Dow. 2021. Framing Creative Work: Helping Novices Frame Better Problems through Interactive Scaffolding. In *Creativity and Cognition*. 1–10. https://doi.org/10.1145/3450741.3465261
[73] Yaoli Mao, Janet Rafner, Yi Wang, and Jacob Sherson. 2023. A hybrid intelligence approach to training generative design assistants: partnership between human experts and AI enhanced co-creative tools. In *HHAI 2023: Augmenting Human Intellect*. IOS Press, 108–123. https://doi.org/10.3233/FAIA230078
[74] Jennifer Marlow and Laura Dabbish. 2014. From rookie to all-star: professional development in a graphic design social networking site. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. 922–933. https://doi.org/10.1145/2531602.2531651
[75] Tuuli Mattelmäki et al. 2006. *Design probes*. Aalto University.
[76] David Maulsby, Saul Greenberg, and Richard Mander. 1993. Prototyping an intelligent agent through Wizard of Oz. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 277–284. https://doi.org/10.1145/169059.169215
[77] Thomas P Moran and John M Carroll. 2020. *Design rationale: Concepts, techniques, and use*. CRC Press.
[78] Roxana Moreno. 2004. Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. *Instructional science* 32, 1 (2004), 99–113. https://doi.org/10.1023/B:TRUC.0000021811.66966.1d
[79] Tricia J Ngoon, C Ailie Fraser, Ariel S Weingarten, Mira Dontcheva, and Scott Klemmer. 2018. Interactive guidance techniques for improving creative feedback. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–11. https://doi.org/10.1145/3173574.3173629
[80] Tricia J Ngoon, Joy O Kim, and Scott Klemmer. 2021. Shöwn: Adaptive Conceptual Guidance Aids Example Use in Creative Tasks. In *Designing Interactive Systems Conference 2021*. 1834–1845. https://doi.org/10.1145/3461778.3462072
[81] Thi Thao Duyen T Nguyen, Thomas Garncarz, Felicia Ng, Laura A Dabbish, and Steven P Dow. 2017. Fruitful Feedback: Positive affective language and source anonymity improve critique reception and work outcomes. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*. 1024–1034. https://doi.org/10.1145/2998181.2998319
[82] David J Nicol and Debra Macfarlane-Dick. 2006. Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in higher education* 31, 2 (2006), 199–218.
[83] Don Norman. 2004. Affordances and design. *Unpublished article, available online at: http://www. jnd. org/dn. mss/affordances-and-design. html* (2004).
[84] Peter O'Donovan, Aseem Agarwala, and Aaron Hertzmann. 2015. Designscape: Design with interactive layout suggestions. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 1221–1224. https://doi.org/10.1145/2702123.2702149
[85] Changhoon Oh, Jungwoo Song, Jinhan Choi, Seonghyeon Kim, Sungwoo Lee, and Bongwon Suh. 2018. I lead, you help but only with enough details: Understanding user experience of co-creation with artificial intelligence. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13. https://doi.org/10.1145/3173574.3174223
[86] Jon L Pierce, Tatiana Kostova, and Kurt T Dirks. 2001. Toward a theory of psychological ownership in organizations. *Academy of management review* 26, 2 (2001), 298–310.

[87] Garr Reynolds. 2011. *Presentation Zen: Simple ideas on presentation design and delivery.* New Riders.

[88] Kathleen Dudden Rowlands. 2007. Check it out! Using checklists to support student learning. *English Journal* (2007), 61–66.

[89] D Royce Sadler. 1989. Formative assessment and the design of instructional systems. *Instructional science* 18, 2 (1989), 119–144. https://doi.org/10.1007/BF00117714

[90] Arvind Satyanarayan and Graham M. Jones. 2024. Intelligence as Agency: Evaluating the Capacity of Generative AI to Empower or Constrain Human Action. *An MIT Exploration of Generative AI* (mar 27 2024). https://mit-genai.pubpub.org/pub/94y6e0f8.

[91] Harmen Schaap, Liesbeth Baartman, and Elly De Bruijn. 2012. Students' learning processes during school-based learning and workplace learning in vocational education: A review. *Vocations and learning* 5 (2012), 99–117. https://doi.org/10.1007/s12186-011-9069-2

[92] Roger C Schank, Tamara R Berman, and Kimberli A Macpherson. 1999. Learning by doing. *Instructional-design theories and models: A new paradigm of instructional theory* 2, 2 (1999), 161–181. https://doi.org/10.1016/0364-0213(94)90007-8

[93] A Schön, Donald. 1984. *The reflective practitioner: How professionals think in action.* Basic Books.

[94] Moushumi Sharmin and Brian P Bailey. 2013. ReflectionSpace: an interactive visualization tool for supporting reflection-on-action in design. In *Proceedings of the 9th ACM Conference on Creativity & Cognition.* 83–92. https://doi.org/10.1145/2466627.2466645

[95] Ben Shneiderman. 2007. Creativity support tools: accelerating discovery and innovation. *Commun. ACM* 50, 12 (2007), 20–32. https://doi.org/10.1145/1323688.1323689

[96] Masaki Suwa and Barbara Tversky. 2002. External representations contribute to the dynamic construction of ideas. In *Diagrammatic Representation and Inference: Second International Conference, Diagrams 2002 Callaway Gardens, GA, USA, April 18–20, 2002 Proceedings 2.* Springer, 341–343. https://doi.org/10.1007/3-540-46037-3_33

[97] Michael Terry, Elizabeth D Mynatt, Kumiyo Nakakoji, and Yasuhiro Yamamoto. 2004. Variation in element and action: supporting simultaneous development of alternative solutions. In *Proceedings of the SIGCHI conference on Human factors in computing systems.* 711–718. https://doi.org/10.1145/985692.985782

[98] Kashyap Todi, Daryl Weir, and Antti Oulasvirta. 2016. Sketchplore: Sketch and explore with a layout optimiser. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems.* 543–555. https://doi.org/10.1145/2901790.2901817

[99] Helena Vasconcelos, Matthew Jörke, Madeleine Grunde-McLaughlin, Tobias Gerstenberg, Michael Bernstein, and Ranjay Krishna. 2022. Explanations Can Reduce Overreliance on AI Systems During Decision-Making. *arXiv preprint arXiv:2212.06823* (2022).

[100] Jeremy Warner, Shuyao Zhou, and Björn Hartmann. 2023. Interactively Optimizing Layout Transfer for Vector Graphics. In *Proceedings of the AI & HCI Workshop at the 40th International Conference on Machine Learning (ICML).*

[101] Robin Williams. 2015. *The non-designer's design book: Design and typographic principles for the visual novice.* Pearson Education.

[102] Naomi E. Winstone, Robert A. Nash, James Rowntree, and Michael Parker. 2017. 'It'd be useful, but I wouldn't use it': barriers to university students' feedback seeking and recipience. *Studies in Higher Education* 42, 11 (2017), 2026–2041. https://doi.org/10.1080/03075079.2015.1130032 arXiv:https://doi.org/10.1080/03075079.2015.1130032

[103] Jun Xiao, John Stasko, and Richard Catrambone. 2004. An Empirical Study of the Effect of Agent Competence on User Performance and Perception. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems - Volume 1* (New York, New York) *(AAMAS '04)*. IEEE Computer Society, USA, 178–185.

[104] Anbang Xu, Shih-Wen Huang, and Brian Bailey. 2014. Voyant: generating structured feedback on visual designs using a crowd of non-experts. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing.* 1433–1444. https://doi.org/10.1145/2531602.2531604

[105] Kenta Yamamoto, Yuki Koyama, and Yoichi Ochiai. 2022. Photographic Lighting Design with Photographer-in-the-Loop Bayesian Optimization. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology.* 1–11. https://doi.org/10.1145/3526113.3545690

[106] Yu-Chun Yen, Steven P Dow, Elizabeth Gerber, and Brian P Bailey. 2016. Social network, web forum, or task market? Comparing different crowd genres for design feedback exchange. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems.* 773–784. https://doi.org/10.1145/2901790.2901820

[107] Yu-Chun Grace Yen, Steven P Dow, Elizabeth Gerber, and Brian P Bailey. 2017. Listen to others, listen to yourself: Combining feedback review and reflection to improve iterative design. In *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition.* 158–170. https://doi.org/10.1145/3059454.3059468

[108] Yu-Chun Grace Yen, Joy O Kim, and Brian P Bailey. 2020. Decipher: an interactive visualization tool for interpreting unstructured design feedback from multiple providers. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* 1–13. https://doi.org/10.1145/3313831.3376380

[109] Alvin Yuan, Kurt Luther, Markus Krause, Sophie Isabel Vennix, Steven P Dow, and Bjorn Hartmann. 2016. Almost an expert: The effects of rubrics and expertise on perceived value of crowdsourced design critiques. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing.* 1005–1017. https://doi.org/10.1145/2818048.2819953

A APPENDIX

principle	issue	specific instance
hierarchy	weak point of entry	The title does not seem particularly emphasized compared to [other text], making it appear only slightly more important.
	ambiguous levels of importance	Relative importance of [element] to [element]/amongst [elements] seems unclear.
	unclear grouping of content	[element] and [element], which seem to convey similar meanings, are not placed within the same area.
alignment	arbitrary alignment of elements	[element] and [element], which seem to be grouped, are not aligned to the same [line/edge/axis].
	insufficient margins	[element] is placed quite close to the [right/left/top/bottom] margin, which can make the design appear somewhat crowded.
balance	not enough space between content	[element] and [element] appear to be quite close to each other, which can make the design appear crowded.
	content lacks balance	Most of your [elements] are in [area], which may make your design appear [direction]-heavy.
	uneven margins	The [left/right or top/bottom] margin of your design appears to be [smaller/larger] than the others [within the frame/design element], which can appear inconsistent or imbalanced.
unity	inconsistent/too many variations in text	Your design uses [number] different [typefaces, font sizes, etc.], which can make your design seem incohesive.
	unnecessary design elements	There seems to be several [graphic] in your design, which can somewhat distract from the main message/theme.
	inconsistent color choices	Your design uses quite a few colors, many of which don't seem to be reflected more than once, which can make your design seem incohesive.
readability	poor text legibility	The [font, line break, etc.] used for/in the [text] makes it somewhat harder to read.
	unsuitable image manipulation	The [contrast/brightness] of [graphic] is quite [low/high], which may make the content somewhat hard to see.
	obscured content	[Element] is [slightly/completely] obscured by [element], which can make it somewhat harder to [read/see] clearly.

Table 2: Our probe implements a feedback framework based on these design principles and specific issues. In our framework, each principle includes a high-level definition, including how it can be achieved in practice. For each issue, we provide an example(s) of customizable "specific instance" text template(s) that describes in detail how the issue is surfaced in the current design. Brackets indicate text to customize.