EXTENDED-ABSTRACT

Feedstack: Layering Structured Representations Over Unstructured Feedback to Scaffold Human–AI Conversation

HANNAH VY NGUYEN, Temple University, Philadelphia, PA, United States

GRACE YU CHUN YEN, University of California, San Diego, San Diego, CA, United States

OMAR SHAKIR, Temple University, Philadelphia, PA, United States

HANG HUYNH, Temple University, Philadelphia, PA, United States

SEBASTIAN GUTIERREZ, Temple University, Philadelphia, PA, United States

JUNE A SMITH, Berea College, Berea, KY, United States

View all

Feedstack: Layering Structured Representations over Unstructured Feedback to Scaffold Human–AI Conversation

Hannah Vy Nguyen
Computer and Information Sciences
Temple University
Philadelphia, Pennsylvania, USA
hannahnguyen4@sigchi.org

Grace Yu-Chun Yen
Cognitive Science
University of California, San Diego
San Diego, California, USA
yyen@ucsd.edu

Omar Shakir
Computer and Information Sciences
Temple University
Philadelphia, Pennsylvania, USA
omar.shakir@temple.edu

Hang Huynh
Computer and Information Sciences
Temple University
Philadelphia, Pennsylvania, USA
hang.huynh@temple.edu

Sebastian Gutierrez
Computer and Information Sciences
Temple University
Philadelphia, Pennsylvania, USA
guts@temple.edu

June A Smith
Computer Science
Berea College
Berea, Kentucky, USA
smithj7@berea.edu

Sheila Jimenez
Computer and Information Sciences
Temple University
Philadelphia, Pennsylvania, USA
sheila.jimenez@temple.edu

Salma E Abdelgelil
Computer and Information Sciences
Temple University
Philadelphia, Pennsylvania, USA
salmaabdelgelil@temple.edu

Stephen MacNeil
Computer and Information Sciences
Temple University
Philadelphia, Pennsylvania, USA
stephen.macneil@temple.edu

Abstract

Many conversational user interfaces facilitate linear conversations with turn-based dialogue, similar to face-to-face conversations between people. However, digital conversations can afford more than simple back-and-forth; they can be layered with interaction techniques and structured representations that scaffold exploration, reflection, and shared understanding between users and AI systems. We introduce Feedstack, a speculative interface that augments feedback conversations with layered affordances for organizing, navigating, and externalizing feedback. These layered structures serve as a shared representation of the conversation that can surface user intent and reveal underlying design principles. This work represents an early exploration of this vision using a research-through-design approach. We describe system features and design rationale, and present insights from two formative (n=8, n=8) studies to examine how novice designers engage with these layered supports. Rather than presenting a conclusive evaluation, we reflect on Feedstack as a design probe that opens up new directions for conversational feedback systems.

CCS Concepts

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

Keywords

feedback, design, sense-making, conversational user interface

1 Introduction

The adoption of conversational user interfaces (CUIs) in education is rapidly growing, especially as students increasingly rely on these tools as a help resource [16, 17] and for feedback [33, 35]. This popularity reflects their accessibility and ability to provide immediate responses, often mimicking the back-and-forth flow of human dialogue. However, they are primarily designed to support linear conversations [19], with few opportunities for divergence, reflection, or revisiting earlier points. As a result, conversations often remain confined to a single, linear path.

In contrast, a few information exploration tools are emerging that prioritize structure over conversation. These diagrammatic systems, such as CausalMapper or Sensecape [18, 31], offer affordances for exploring connections between ideas, but they sacrifice the fluid, iterative nature of dialogue. This demonstrates a current gap in design approaches, which we explore as the following research question:

RQ: How might we adapt conversational user interfaces to better support exploration and reflection while maintaining the unstructured nature of dialogue?

Some chatbots currently support exploration by mentioning related topics or offering follow-up questions; however, the burden is still placed on the user to notice connections and pursue them. This limitation is especially problematic for non-expert users, who may struggle to articulate what they need or to understand the rationale

behind the feedback they receive [8]. Without conversational cues or scaffolds that encourage reflection and exploration, users may accept feedback passively, without internalizing it, understanding it, or relating it to their broader goals.

In this paper, we investigate the potential to layer structured representations to support exploration and reflection. These 'shared representations also serve to scaffold human-AI interaction, where implicit aspects of the conversation become explicit. To investigate this potential, we introduce **Feedstack**, a multimodal conversational user interface that actively supports exploration and reflection, while also maintaining the natural flow of the conversation. This design probe [3] was developed based on design insights from a formative study. Through a second user study, we identify additional opportunities for refinement, which have been integrated into a web-based prototype. This research-through-design approach [40] is intended to advance our vision for chatbots that are anchored to shared representations that externalize information that is often implicit within the conversation or slightly outside the conversational context.

2 Related Work

2.1 The Nature of Feedback Conversations

Feedback is an essential component of effective learning, especially in domains like design, where iterative improvement is a critical skill to develop and students learn through critique from experts and peers [37]. However, students often struggle to interpret and apply unstructured feedback. Students also face challenges in knowing which parts of the feedback to prioritize [2] and how specific instances of the feedback relate to the underlying design principles [12]. Previous research has shown that structured feedback (i.e., feedback that is clearly organized and tied to specific learning objectives) improves comprehension and application, which has led to numerous intelligent creativity support tools that focus on structuring unstructured feedback [9, 36].

Feedback conversations are highly interactive and can meander into unexpected topics and critiques at any point during the conversation. For example, when an expert provides feedback, a student might respond with a clarifying question, especially when the feedback introduces unfamiliar topics or concepts. Due to the 'curse of expertise' [14], these disconnects between the ideas being expressed by experts and how they are interpreted by novices are common.

One approach to help students interpret feedback is to connect feedback instances to high-level design principles [15]. Prior work suggests that this can help students understand the broader context and underlying justifications behind the feedback [5]. Effective feedback also encourages students to reflect on how their work fits into 'the big picture' by helping learners see their work from various perspectives and see connections in the feedback. Unfortunately, in unstructured feedback, these connections are often implicit. We explore how mapping and externalizing high-level design principles in feedback can enhance student learning, particularly in chatbot conversations, by helping students better connect feedback to the broader design context and improve their decision-making.

2.2 Structuring Unstructured Design Feedback

Chatbots are increasingly used in educational settings to support teachers and students [24, 28, 30]. Many students now turning to general purpose chatbots, such as ChatGPT, to answer their questions, explain content, and guide their learning [16, 17]. These interfaces are familiar and largely replicate some of the natural language interactions they might have with peers and instructors. Recently, the performance of these chatbots has improved dramatically, requiring fewer explicit rules and less fine-tuning. The release of multimodal models now enables users to incorporate visual imagery in their chatbot conversations;for example, visual question answering [22].

However, research on intelligent feedback systems has consistently emphasized the importance of structuring feedback to improve learning and sense-making [1, 21, 25, 26, 34, 36, 39]. For example, Sensecape, CausalMapper, and Graphologue address the challenge of information exploration by providing multi-level abstractions and interactive diagrams [18, 20, 32]. Similarly, Decipher and Voyant provide feedback scaffolding by structuring feedback based on topics and intentions to help learners act on feedback more effectively [36, 38].

We have identified a research gap in designing chatbots that support fluid feedback conversations, in a way that enables users to layer structured insights on top of unstructured conversations. Specifically, there is a need for CUIs that maintain the natural flow of feedback while also encouraging students to explore feedback topics concurrently, rather than in a strictly linear manner.

3 Formative Study

To explore how novice designers engage with traditional CUIs in a design feedback context, we conducted a formative study. Our study investigated how standard chatbot interfaces (i.e. ChatGPT) facilitates or fails to facilitate feedback interactions for non-expert users. Participants were tasked with engaging in a feedback conversation with ChatGPT to stimulate a design critique session. Designers increasingly interact with chat interfaces that are integrated within their design tools (e.g.: Figma); however, these chat interfaces are designed to support design activity directly, and not to provide feedback. Designers often still rely on external feedback sources such as a peer, advisor, or ChatGPT.

We recruited 8 university-level students from various institutions, including 6 undergraduates and 2 graduates from diverse majors, such as Chemistry, Education, Computer Science, and Psychology. Participants were recruited via university mailing lists and communication channels, including Slack. Each participant received $20 USD compensation in the form of a digital gift card for their time. We did not require any prior experience with formal feedback systems or ChatGPT. Participants were provided with a fictional website mockup and instructed to imagine themselves as beginner designers revising their interface based on expert feedback. Participants could freely interact with ChatGPT and the sessions lasted 7-10 minutes. The team qualitatively coded the conversations to understand the patterns and dynamics of how the participants interacted with the ChatGPT interface. The study was approved by IRB.
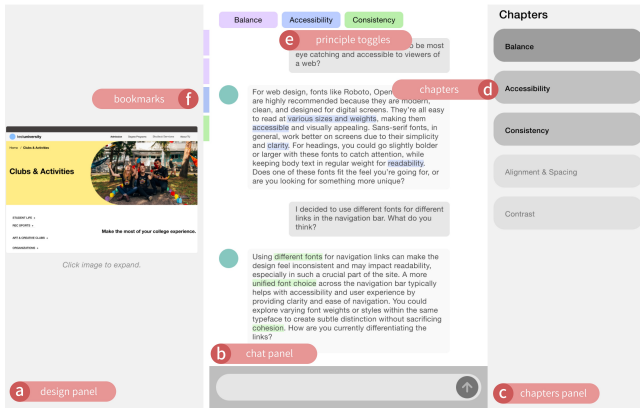
Figure 1: The design panel A displays the user's design artifact. The chat panel B contains the chat where the user can interact with the LLM-powered design expert to receive and explore feedback. The chapters panel C contains the interactive chapters D. On the top of the chat panel is the principle toggles E and on the left of the chat panel is the bookmarks F.

Our main findings suggest that users use ChatGPT for requesting examples and asking clarifying questions. They sometimes had trouble leading the conversation and other times changed topics to guide their exploration of design topics. Based on these findings and the existing literature, we devised the following design goals:

> **D1. Preserve the Open-Ended Nature of Feedback** Retain the unstructured, conversational quality of feedback discussions to support ambiguity, nuance, and interpretation.

> **D2. Support Exploration** Enable users to connect individual pieces of feedback to broader design principles. Encourage discovery of relevant but undiscussed principles through prompts, suggestions, or visualizations that surface thematic gaps.

> **D3. Promote Reflection** Help users track what has been discussed (and what has not) across the conversation. Surface patterns of topics or absences to promote reflection.

4 Feedstack Design

We designed an interactive system called **Feedstack**, an interface that scaffolds real-time feedback conversations between the user and the LLM through interactions and affordances outlined below. The name relates to the idea that structured feedback is *'stacked'* on top of an unstructured feedback conversation, kind of like how an integrated development environment (IDE) provides affordances around the code. The chatbot interface is in the center (See B in Figure 1) with peripheral panels on the left and right. This conversation revolves around a visual design, which users can upload and view through the conversation (See A in Figure 1).

4.1 Bookmarks

The Bookmarks feature (See F in Figure 1) visually marks points along the discussion's scrub bar where specific design principles were discussed. By showing these segments, it allows users to quickly locate and revisit moments when particular principles were discussed, similar to how 'Find' features in PDF viewers highlight all occurrences of a keyword in a scroll bar.

4.1.1 Design Rationale. This design supports *reflection* [29] by allowing novice designers to easily revisit key parts of the conversation, enabling focused review and deeper thinking about each principle. It also reveals the extent to which each principle was discussed, highlighting potential over- or under-emphasis. The design also aligns with *variation theory* [11, 27] by helping users quickly navigate to multiple examples of how a given principle was applied. Juxtaposing these instances allows learners to notice subtle differences in interpretation and application across contexts.

4.2 Chapters

As design principles are identified in the conversation, they appear on the side panel as Chapters (See D) to help make these implicitly discussed design principles explicit to users. Chapters are generated in real time using a large language model to analyze the conversation. In our initial prototype, the chapters were represented as five pre-defined accordions, each representing a core principle derived from established visual design principles [23]. Chapters also include the following features:

- *Learning Materials:* Each expanded chapter is organized into three sections: 1) *Principle Definition:* A clear explanation of the principle, 2) *Relation to Your Design*: Insights about how the principle applies to the user's design, 3) *Key Terms*: A glossary of important terminology related to the principle. These materials are generated using a large language model when the design principle is first identified in the conversation.
- *Collapsibility:* To minimize cognitive load, all chapters are initially presented in a collapsed state. This design reduces the visible information, helping users focus on specific topics without feeling overwhelmed. Users can open or close the accordions at any time.
- *Opacity* As users interact with each principle during feedback conversations, the opacity of the corresponding chapter gradually increases. This increase in opacity reflects the frequency with which the principle is referenced, signaling to users and the chatbot which topics are of interest and which are underexplored.

4.2.1 Design Rationale. The Chapters were designed as a shared representation between the AI and user. In mixed-initiative systems, *shared representations* [13] can provide common ground between the agent and the user, facilitating mutual understanding. In the context of feedback conversations, the Chapters externalize the design principles that often remain tacit or implicit within natural language exchanges. This externalization helps users zoom out from the back-and-forth flow of conversation and see the larger structure and themes that are unfolding.
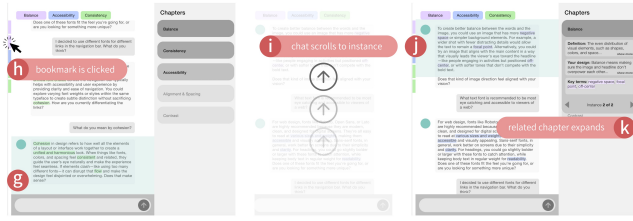
Figure 2: The chat initially anchors at the bottom of the panel G to the most recent message. When the user selects a bookmark H, such as the 'Balance' bookmark, the chat automatically scrolls upwards I to the most recent instance where 'Balance' is discussed in the feedback J. At the same time, the corresponding chapter for 'Balance' expands in the sidebar K.

4.3 Highlights

Key terms related design principles are highlighted to focus the user's attention on important areas of feedback. As new design principles emerge in the discussion, they are listed at the top of the chat conversation (see E in Figure 1). By clicking on them, users can turn the highlights on, or turn them off to reduce their cognitive load.

4.3.1 Design Rationale. Drawing on dual-coding theory [7], the highlights augment text-based feedback with visual cues to help users better understand which design principles are being emphasized during the conversation.

5 User Study

5.1 Methodology

To further develop our vision for layered scaffolding, we conducted a user study with design novices to understand whether and how these features might support the design goals. Given the exploratory nature of this work, we follow a research-through-design method where the focus is on developing design insights as opposed to falsifying hypotheses [40]. Participants in the study were asked to think aloud during the study and their responses were analyzed through reflexive thematic analysis [4]. The study was approved by IRB.

Study Design. We designed a fictional conversation using the real user utterances observed in the formative study. We populated the conversation as text in the prototype, enabling users to progress through the design feedback conversation while interacting with the system features.

Participants and Recruitment. We recruited 8 participants (5 female, 2 male, 1 non-binary) from North American colleges and universities. Participants self-identified as beginner designers in the pre-study survey. The exclusion criteria included prior completion of an intermediate-level design course or professional experience as a designer. Participants received a $20 USD digital gift card upon completing the study.

Procedure. Participants began by completing an informed consent and pre-study survey, followed by analyzing a fictional website mockup and identifying areas for improvement based on visual

design principles. They were introduced to five core design principles: *Accessibility, Consistency, Contrast, Balance,* and *Alignment and Spacing*. These principles were explained to help them recognize potential areas for improvement within their designs. After a brief tutorial on navigating the prototype, they engaged with the system while thinking aloud (15 minutes), progressing through the mock conversation and interacting with system features. The session concluded with a post-study survey and semi-structured interview, where participants reflected on their experience.

5.2 Results

5.2.1 Reflections on Learning and Design Principles. Multiple participants described how their understanding of the design principles improved as a result of using Feedstack. Several participants reported that they began to notice principles they had previously overlooked or not considered relevant. For instance, P6 said:

> *"Consistency surprised me. It wasn't something that stood out at first, but now I see how it applies. Also, under balance, things like negative space came up, which I didn't think of initially."*

Similarly, P4 shared, *"I didn't notice the balance issue with the image at first. I thought it was fine, but when feedback mentioned it being dominant, it made sense."* These responses suggest that the system helped surface implicit design principles, encouraging deeper reflection on their applicability.

Participants expressed the value of learning design principles when receiving feedback. For example, P2 shared,

> *"Learning about the principles helps change how you approach designing, while specific feedback is more about fixing something. It's like the analogy—if you give someone a fish, they eat for a day, but if you teach them to fish, they can feed themselves. Learning the design principles is just as important as feedback."*

Additionally, they shared, *"I think I've overlooked accessibility in design. I usually focus more on aesthetics and less on the practical side of things, so it was interesting to see accessibility as a core consideration."*

5.2.2 Desire to Explore Deeper. Beyond recognizing principles, participants expressed interest in exploring the key terms and highlights. For example, P7 noticed multiple instances of 'Balance' throughout the feedback conversation:

> *"I just noticed the other instances of balance that are highlighted. It's nice to see how the feedback ties into the principles."*

Similarly, P4 attempted to click on the highlights, expecting a definition. They shared,

> *"I liked the key terms, though I wished I could click on them to get definitions directly from the paragraph. That's just what I kept trying to do, even though the definitions were already provided [within the feedback]."*

This indicates that the user is motivated to engage with the terms at a more granular level, highlighting the potential for the tool to go beyond providing feedback and serve as an interactive learning environment, where students can independently explore and clarify design concepts and connect them to underlying design principles.

5.2.3 Continued Use After the Conversation. Several participants engaged with the system's Chapters after the conversation concluded, engaging with the affordances to reflect on the feedback and further explore the design principles post-conversation. P7 noted, *"It's useful for summarizing, but also for quickly referencing the conversation in more detail. If I want to remember something specific, it's helpful to have the full context behind it."* This suggests that Feedstack could enable users to revisit feedback at later stages, reinforcing learning and aiding in the iterative design process.

In the context of iteration, participants such as P8 expressed appreciation for the system's navigational features, sharing that they would use them more after the conversation to implement changes. P8 shared,

> *"It's great for navigating through the conversation without scrolling. I'd probably use this feature more after the conversation when I'm implementing changes."*

This indicates that Feedstack's interactive feedback structure is not only useful for the conversation itself but also supports users in applying feedback and refining their designs.

6 Discussion

6.1 Synthesizing Design Insights

6.1.1 Reflection on Design Principles. The results of the user study suggest that Feedstack helps novice designers become more aware of design principles and engage with them more deeply. Participants also seemed to be more appreciative of the design principles beyond just the design feedback. These findings are promising and suggest that the provided affordances have the potential to make the implicit aspects of the conversation explicit.

6.1.2 Support for Exploration and Reflection. Feedstack appears to encourage active exploration of key terms and design principles. Participants expressed interest in exploring the highlighted key terms more extensively. They also described how Feedstack encouraged them to reflect on the feedback even after the conversation 'ended.' This suggests that the features are helping students go beyond the conversation and engage more holistically with the feedback.

6.2 Implications for Future Design

The feedback from participants also points to areas for improvement. For instance, offering more interactive features—such as clickable key terms for direct definitions—could improve users' experience. This could make the tool even more effective in supporting users' learning and promote exploration.

During our user studies, the main issue we observed was challenges in keeping the conversation going. This inspired us to present two additional features which have been integrated into our functioning prototype. In addition to these features, we also added a feature for referencing feedback instances. The most recent prototype which is implemented in React.js, Django, and uses the OpenAI API is shown in Figure 3.

6.2.1 Emerging Topics. One design idea that emerged from this work was to extend the opacity feature to show suggested topics that are related but have not yet been introduced to the conversation.
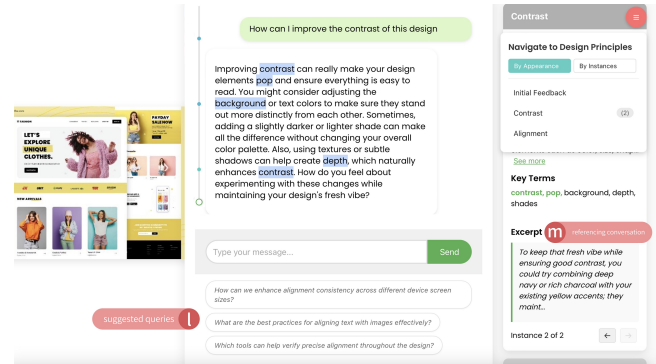

Figure 3: The resulting web-based Feedstack system. It incorporates new features informed by recent findings. These include L suggested queries and M excerpt references that highlight the principle being applied during the conversation.

This provides information scents [6] towards concepts and ideas that are just slightly beyond the immediate conversational context.

6.2.2 Conversational Cues L. Chatbots can offer conversational cues, such as suggesting conversational turns for the user that they may want to take [10]. This approach could be helpful to guide users toward new topics and ideas. Compared with the *Emerging Topics* feature, we expect this to be more oriented toward feedback instances (clarifying feedback or making it more actionable) as opposed to high-level design principles.

6.2.3 Referencing Conversation M. Participants explained that they wanted more opportunities to see instances of design feedback for specific design principles. This inspired a navigation area within the Chapters with references to feedback instances in the conversation.

6.3 Limitations and Future Work

Our work is rooted in the research-through-design tradition, where the goal is not to validate hypotheses or to offer conclusive evidence. Rather, the goal is to explore and also to communicate design ideas. The prototype in this study serves as a design probe [3] to instantiate abstract ideas. However, as future work, the prototype should be evaluated more explicitly with comparisons to existing commercial chatbots.

7 Conclusion

Our research introduces a proof-of-concept interactive system, Feedstack. We explore the use of novel affordances and interactions to structure design feedback conversations, externalizing underlying principles and encouraging users to more deeply engage with them beyond the primary conversation. Our preliminary evaluation indicates that novice designers found value in layered affordances that promote awareness of previously hidden principles.

Acknowledgments

References

[1] Salvatore Andolina, Hendrik Schneider, Joel Chan, Khalil Klouche, Giulio Jacucci, and Steven Dow. 2017. Crowdboard: augmenting in-person idea generation with real-time crowds. In *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition*. 106–118.

[2] Alasdair Blair, Steven Curtis, Mark Goodwin, and Sam Shields. 2013. What Feedback do Students Want? *Politics* 33, 1 (Feb. 2013), 66–79. https://doi.org/10.1111/j.1467-9256.2012.01446.x

[3] Kirsten Boehner, William Gaver, and Andy Boucher. 2012. Probes. In *Inventive methods*. Routledge, 185–201.

[4] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative research in sport, exercise and health* 11 (2019).

[5] Joel Chan, Zijian Ding, Eesh Kamrah, and Mark Fuge. 2024. Formulating or Fixating: Effects of Examples on Problem Solving Vary as a Function of Example Presentation Interface Design. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–16. https://doi.org/10.1145/3613904.3642653

[6] Ed H Chi, Peter Pirolli, Kim Chen, and James Pitkow. 2001. Using information scent to model user information needs and actions and the Web. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 490–497.

[7] Jim Clark and Allan Paivio. 1991. Dual Coding Theory and Education. *Educational Psychology Review* 3 (09 1991), 149–210. https://doi.org/10.1007/BF01320076

[8] Eureka Foong, Darren Gergle, and Elizabeth M. Gerber. 2017. Novice and Expert Sensemaking of Crowdsourced Design Feedback. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 45 (Dec. 2017), 18 pages. https://doi.org/10.1145/3134680

[9] C Ailie Fraser, Tricia J Ngoon, Ariel S Weingarten, Mira Dontcheva, and Scott Klemmer. 2017. CritiqueKit: A mixed-initiative, real-time interface for improving feedback. In *Adjunct Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*. 7–9.

[10] Marcel Gohsen, Johannes Kiesel, Mariam Korashi, Jan Ehlers, and Benno Stein. 2023. Guiding Oral Conversations: How to Nudge Users Towards Asking Questions?. In *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval* (Austin, TX, USA) (CHIIR '23). Association for Computing Machinery, New York, NY, USA, 34–42. https://doi.org/10.1145/3576840.3578291

[11] Lingyuan Gu, Rongjin Hunag, and Ference Marton. 2004. Teaching with variation: A Chinese way of promoting effective mathematics learning. In *How Chinese learn mathematics: Perspectives from insiders*. World Scientific, 309–347.

[12] John Hattie and Helen Timperley. 2007. The Power of Feedback. *Review of Educational Research* 77, 1 (March 2007), 81–112.

[13] Jeffrey Heer. 2019. Agency plus automation: Designing artificial intelligence into interactive systems. *Proceedings of the National Academy of Sciences* 116, 6 (2019), 1844–1850.

[14] Pamela J Hinds. 1999. The curse of expertise: The effects of expertise and debiasing methods on prediction of novice performance. *Journal of experimental psychology: applied* 5, 2 (1999), 205.

[15] Pamela J. Hinds, Michael Patterson, and Jeffrey Pfeffer. 2001. Bothered by abstraction: The effect of expertise on knowledge transfer and subsequent novice performance. *Journal of Applied Psychology* 86, 6 (2001), 1232–1243. https://doi.org/10.1037/0021-9010.86.6.1232

[16] Irene Hou, Sophia Mettille, Owen Man, Zhuo Li, Cynthia Zastudil, and Stephen MacNeil. 2024. The Effects of Generative AI on Computing Students' Help-Seeking Preferences. In *Proceedings of the 26th Australasian Computing Education Conference* (Sydney, NSW, Australia) (ACE '24). Association for Computing Machinery, New York, NY, USA, 39–48. https://doi.org/10.1145/3636243.3636248

[17] Irene Hou, Hannah Vy Nguyen, Owen Man, and Stephen MacNeil. 2025. The Evolving Usage of GenAI by Computing Students. In *Proceedings of the 56th ACM Technical Symposium on Computer Science Education V. 2 (SIGCSE TS 2025)*. ACM, 1481–1482. https://doi.org/10.1145/3641555.3705266

[18] Ziheng Huang, Kexin Quan, Joel Chan, and Stephen MacNeil. 2023. CausalMapper: Challenging designers to think in systems with Causal Maps and Large Language Model. In *Proceedings of the 15th Conference on Creativity and Cognition*. 325–329.

[19] Mohit Jain, Pratyush Kumar, Ramachandra Kota, and Shwetak N. Patel. 2018. Evaluating and Informing the Design of Chatbots. In *Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18)*. ACM, 12 pages. https://doi.org/10.1145/3196709.3196735

[20] Peiling Jiang, Jude Rayan, Steven P. Dow, and Haijun Xia. 2023. Graphologue: Exploring Large Language Model Responses with Interactive Diagrams. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–20. https://doi.org/10.1145/3586183.3606737

[21] 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*.

[22] Gyeong-Geon Lee and Xiaoming Zhai. 2024. Realizing Visual Question Answering for Education: GPT-4V as a Multimodal AI. *ArXiv* abs/2405.07163 (2024).

https://api.semanticscholar.org/CorpusID:269757576

[23] William Lidwell, Kritina Holden, and Jill Butler. 2020. *Universal Principles of Design* (3rd ed.). Rockport Publishers, Beverly, MA.

[24] Li Liu, Rama Subbareddy, and C. G. Raghavendra. 2022. AI Intelligence Chatbot to Improve Students Learning in the Higher Education Platform. *Journal of Interconnection Networks* 22 (2022). https://doi.org/10.1142/S0219265921430325

[25] Kurt Luther, Amy Pavel, Wei Wu, Jari-lee Tolentino, Maneesh Agrawala, Björn Hartmann, and Steven P Dow. 2014. CrowdCrit: crowdsourcing and aggregating visual design critique. In *Proceedings of the companion publication of the 17th ACM conference on Computer supported cooperative work & social computing*. Association for Computing Machinery, New York, NY, USA, 21–24.

[26] 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*. Association for Computing Machinery, New York, NY, USA, 473–485.

[27] Lauren Margulieux, Paul Denny, Kathryn Cunningham, Michael Deutsch, and Benjamin R Shapiro. 2021. When wrong is right: The instructional power of multiple conceptions. In *Proceedings of the 17th ACM Conference on International Computing Education Research*. 184–197.

[28] Mónica De La Roca, Miguel M. Chan, Antonio Garcia-Cabot, Eva Garcia-Lopez, and Héctor Amado-Salvatierra. 2024. The impact of a chatbot working as an assistant in a course for supporting student learning and engagement. *Computer Applications in Engineering Education* 32, 5 (2024), e22750. https://doi.org/10.1002/cae.22750 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/cae.22750

[29] Donald A Schön. 2017. *The reflective practitioner: How professionals think in action*. Routledge.

[30] Stefan Sonderegger and Sabine Seufert. 2022. Chatbot-mediated Learning: Conceptual Framework for the Design of Chatbot Use Cases in Education. In *Proceedings of the 14th International Conference on Computer Supported Education - Volume 1: CSEDU*. INSTICC, SciTePress, 207–215. https://doi.org/10.5220/0010999200003182

[31] Sangho Suh, Bryan Min, Srishti Palani, and Haijun Xia. 2023. Sensecape: Enabling multilevel exploration and sensemaking with large language models. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–18.

[32] Sangho Suh, Bryan Min, Srishti Palani, and Haijun Xia. 2023. Sensecape: Enabling Multilevel Exploration and Sensemaking with Large Language Models. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. ACM, San Francisco CA USA, 1–18. https://doi.org/10.1145/3586183.3606756

[33] Olia Tsivitanidou and Andri Ioannou. 2021. Envisioned Pedagogical Uses of Chatbots in Higher Education and Perceived Benefits and Challenges. In *Learning and Collaboration Technologies: Games and Virtual Environments for Learning*, Panayiotis Zaphiris and Andri Ioannou (Eds.). Vol. 12785. Springer International Publishing, Cham, 230–250. https://doi.org/10.1007/978-3-030-77943-6_15 Series Title: Lecture Notes in Computer Science.

[34] Helen Wauck, Yu-Chun Yen, Wai-Tat Fu, Elizabeth Gerber, Steven P Dow, and Brian P Bailey. 2017. From in the class or in the wild? Peers provide better design feedback than external crowds. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 5580–5591.

[35] Yangyu Xiao and Yuying Zhi. 2023. An Exploratory Study of EFL Learners' Use of ChatGPT for Language Learning Tasks: Experience and Perceptions. *Languages* 8 (09 2023), 212. https://doi.org/10.3390/languages8030212

[36] 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*. ACM, Baltimore Maryland USA, 1433–1444. https://doi.org/10.1145/2531602.2531604

[37] 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*. ACM, Singapore Singapore, 158–170. https://doi.org/10.1145/3059454.3059468

[38] 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.

[39] 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.

[40] John Zimmerman, Jodi Forlizzi, and Shelley Evenson. 2007. Research through design as a method for interaction design research in HCI. In *Proceedings of the SIGCHI Conference on Human factors in Computing Systems*. 493–502.