

UNDERSTANDING CO-CREATIVE PROCESSES THROUGH DIALOGUE ACROSS
HUMAN-HUMAN AND HUMAN-AI PARTNERSHIPS

By

AMANDA E. GRIFFITH

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2025

© 2025 Amanda E. Griffith

To Daniel

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor, Dr. Kristy Boyer, for her guidance, patience, and support throughout my doctoral journey. I am also grateful to my committee members for their valuable feedback and insights that helped shape this research. Special thanks to my colleagues in the LearnDialogue Group for their collaboration and encouragement. Finally, I am deeply thankful to my family and friends for their unwavering support and understanding during this process.

This research was supported by the National Science Foundation (NSF) and the Institute of Education Sciences (IES) under grants DRL-1813740 and DRL-1814083 (2020-2023), and by NSF grant 2229612 (2023-2025). The opinions, findings, and conclusions expressed in this dissertation are mine and do not necessarily reflect the views of the funding agencies.

TABLE OF CONTENTS

	<u>page</u>
ACKNOWLEDGEMENTS	4
LIST OF TABLES.....	10
LIST OF FIGURES.....	12
ABSTRACT.....	13
CHAPTER	
1 INTRODUCTION	15
1.1 Research Motivation	16
1.2 Contributions.....	18
1.3 Overview.....	19
2 LITERATURE REVIEW.....	20
2.1 Background and Motivation	20
2.2 Methods.....	21
2.2.1 Sampling	21
2.2.2 Analysis	23
2.3 Creative Task Support.....	24
2.3.1 Design.....	25
2.3.2 Narrative.....	30
2.3.3 Idea Management	31
2.3.4 Music/Audio	33
2.4 Trends in Task Support.....	34
2.4.1 Design Over Time.....	35
2.4.2 Narrative Over Time.....	37
2.4.3 Implications and Future Directions	40
2.5 Collaboration.....	42
2.5.1 Collaboration Analysis	43
2.5.2 Intelligent System Support.....	43
2.6 Key Takeaways and Research Trajectory	44
3 DISCOVERING CO-CREATIVE DIALOGUE STATES DURING COLLABORATIVE LEARNING	47
3.1 Background and Motivation	47
3.2 System Description	49
3.3 Methods.....	50
3.3.1 Learning Environment for Computational Music Remixing.....	51
3.3.2 Dialogue Tagging	52
3.3.3 Hidden Markov Model.....	52
3.4 Results	53

3.5	Discussion.....	55
3.5.1	Dialogue States	55
3.5.2	The Debugging Process and the Conversation it Inspires.....	57
3.6	Key Takeaways for My Dissertation Research	59
4	INVESTIGATING THE RELATIONSHIP BETWEEN DIALOGUE STATES AND PARTNER SATISFACTION DURING CO-CREATIVE LEARNING TASKS	61
4.1	Background and Motivation	61
4.2	Related Work.....	63
4.3	Methods.....	66
4.3.1	Partner Satisfaction Survey	67
4.3.2	Using Hidden States and Transitions to Predict Combined Satisfaction	68
4.4	HMM's Relationship with Combined Satisfaction	71
4.4.1	Compilation Error State's Negative Relationship with Combined Satisfaction..	71
4.4.2	Curriculum Browsing → Code Editing: Negative Relationship with Combined Satisfaction	76
4.4.3	Aesthetic Dialogue → Technical Dialogue: Negative Relationship with Combined Satisfaction	79
4.5	Implications for Intelligent System Support.....	86
4.5.1	Support for Error Resolution.....	86
4.5.2	Support for Help Seeking	87
4.5.3	Scaffolding Aesthetic and Technical Elements.....	88
4.6	Key Takeaways for My Dissertation Research	89
5	DEVELOPING A CO-CREATIVE AI PARTNER FOR CREATIVE COMPUTING	91
5.1	Background and Motivation	91
5.2	System Description	92
5.2.1	EarSketch Overview.....	92
5.2.2	CAI Overview.....	94
5.2.3	CAI Subsystems.....	94
5.2.4	Integration with EarSketch.....	94
5.3	Human-Human Studies as Foundation for CAI	96
5.3.1	Importance of Formative Project Planning Discussions	96
5.3.2	Balancing Technical and Aesthetic Components.....	97
5.3.3	Proactive Error Support.....	97
5.4	First CAI Prototype	98
5.4.1	System Design	98
5.4.2	Study Design.....	99
5.4.3	Findings	101
5.5	Second CAI Prototype	102
5.5.1	Study Design.....	102
5.5.2	Findings	103
5.6	Wizard of Oz Study.....	105
5.6.1	Study Design.....	105

5.6.2	Findings	106
5.7	Third CAI Prototype	108
5.7.1	System Design	108
5.7.2	Spring 2023: Classroom Studies	109
5.7.3	Findings	110
5.7.4	Summer 2023: Summer Camp & Interviews	111
5.7.5	Findings	111
5.8	Key Takeaways for My Dissertation Research	112
6	INVESTIGATING HUMAN-AI CO-CREATIVITY IN NARRATIVE WRITING.....	114
6.1	Motivation and Research Questions	114
6.2	Study Overview.....	115
6.2.1	Participants and Recruitment.....	115
6.2.2	Procedure.....	117
6.3	System Design and Development	118
6.3.1	Interface Design Process.....	118
6.3.2	Technical Architecture.....	120
6.3.3	User Interface Components	121
6.3.4	Logging and Data Capture	121
6.3.5	Agent Implementation	122
6.4	RQ1 Analysis and Results	124
6.4.1	H1: Text Length.....	124
6.4.2	H2: Linguistic Features.....	126
6.4.3	H3: Referential Cohesion.....	130
6.4.4	H4: Deep Cohesion	132
6.4.5	H5: Semantic Homogeneity.....	134
6.5	RQ2 Analysis and Results	137
6.5.1	RQ2: Mapping Participant-AI Content Attribution	138
6.5.2	RQ2: Dialogue Act Analysis.....	139
6.5.3	RQ2: Questionnaires	144
6.6	Discussion.....	149
6.6.1	Creative Agency and Ownership in Human–AI Collaboration.....	150
6.6.2	Enhancement and Homogenization	152
6.6.3	Implications for Human-AI Co-Creativity.....	154
6.7	Limitations and Future Directions	155
7	CONCLUSION.....	163
7.1	Summary of Research Progression and Findings	163
7.1.1	Studies 1 & 2: Understanding Co-Creativity	163
7.1.2	Study 3: Design and Evaluation of a Co-Creative AI Agent.....	164
7.1.3	Study 4: Impact of AI Assistance on Narrative Writing	165
7.2	Dynamics of Co-Creative Partnership	165
7.2.1	The Centrality of Dialogue in Collaboration	165
7.2.2	Agency, Asymmetry, and Creative Contribution in Human-AI Interaction	166

7.3	Open Questions and Future Research Directions	167
7.3.1	Creative Identity and AI Engagement.....	167
7.3.2	Evolution of AI from Tools to Creative Partners.....	167
7.3.3	The Tradeoff Between Individual Linguistic Enhancement and Collective Homogenization.....	168
7.3.4	Long-term Implications for Human Creative Development.....	169
7.4	Concluding Remarks.....	169
APPENDIX		
A	DIALOGUE ACT TAXONOMY FOR HUMAN-HUMAN CO-CREATIVITY	171
B	HIDDEN MARKOV MODEL DIAGRAM.....	172
C	CAI DESIGN COMPONENTS.....	173
D	WARMUP ACTIVITIES	174
D.1	AI Assisted Warmup	174
D.2	Individual Warmup	174
E	WRITING PROMPTS	175
E.1	AI Assisted Condition Writing Prompt.....	175
E.2	Individual Condition Writing Prompt	175
F	POST-ACTIVITY QUESTIONNAIRE.....	176
G	CREATIVE SELF-EFFICACY AND CREATIVE PERSONAL IDENTITY QUESTIONNAIRE.....	177
H	WRITING ATTITUDES AND WRITING SELF-EFFICACY QUESTIONNAIRE	178
I	LANGUAGE BACKGROUND QUESTIONNAIRE	179
J	HIERARCHICAL CODING SCHEMA FOR REQUESTS IN AI-ASSISTED WRITING	180
K	QUESTIONNAIRE ATTENTION CHECK QUESTIONS	182
K.1	Attention Check Measures.....	182
K.2	Scoring Methodology	183
K.3	Attention Check Results.....	183
LIST OF REFERENCES.....		185
BIOGRAPHICAL SKETCH		205

LIST OF TABLES

<u>Tables</u>	<u>page</u>
2-1 Literature Distribution Across Primary Categories.....	24
2-2 Literature Distribution Across Subcategories	24
3-1 State Transition in Student Collaborative Dialogue	56
3-2 Student Dialogue Sequence Following Code Compilation.....	58
4-1 Predictors of Combined Satisfaction in Co-creative Learning.....	71
4-2 Dialogue Excerpt of Task Completion with Unresolved Compilation Errors	73
4-3 Dialogue Excerpt of Task Completion Without Errors	73
4-4 Dialogue Excerpt of Successful Error Resolution	75
4-5 Dialogue Excerpt of Resolution and Explanation of Code Modifications	76
4-6 Dialogue Excerpt of Unresolved Task Confusion	76
4-7 Interaction Log Excerpt of Targeted Usage of Curriculum Resources	77
4-8 Interaction Log Excerpt of Sequential Curriculum Resource Access Pattern.....	78
4-9 Dialogue Excerpt of Task Prioritization	80
4-10 Dialogue Excerpt of Initial Task Planning	81
4-11 Dialogue Excerpt of Late-Stage Error Resolution Attempts.....	82
4-12 Dialogue Excerpt of Sound Selection Strategy	84
4-13 Dialogue Excerpt of Initial Aesthetic Focus in Task Approach.....	85
5-1 Dialogue Excerpt of Formative Planning Stages	97
5-2 Student Perceptions of CAI Survey Questions	100
6-1 MANOVA Results of Linguistic Features	129
6-2 Univariate Comparisons and Descriptive Statistics of Linguistic Features by Condition .	130
6-3 Type III ANOVA Results for Referential Cohesion Mixed-Effects Model.....	132
A-1 Co-creative Dialogue Act Taxonomy.....	171
C-1 CAI Design Components.....	173
F-1 Post-Activity Evaluation	176
G-1 Creative Self-Efficacy and Creative Personal Identity Questionnaire	177
H-1 Writing Attitudes and Writing Self-Efficacy Questionnaire	178

I-1	Language Background Questionnaire	179
J-1	Primary Dialogue Acts.....	180
J-2	Request Subcode Taxonomy Table	181
K-1	Description of Attention Check Questions.....	182
K-2	Attention Check Response Rates (%) by Question	183

LIST OF FIGURES

<u>Figures</u>	<u>page</u>
2-1 PRISMA Flow Diagram	22
2-2 Interactive drawing tutorial generated by the DrawMyPhoto system	26
2-3 Mixplorer Interface	27
2-4 Morai Maker Level Editor.....	28
2-5 The CreaTable media sequencer system	29
2-6 The We-toon system	31
2-7 CreativeBot Interaction Flow	32
2-8 SemanticCollage Interface	33
2-9 The CoExploerer Interface	34
2-10 Temporal Trends in Human-AI Co-Creation Research.....	35
2-11 Distribution of research focus in Design subcategories over time.....	36
2-12 Distribution of research focus in Narrative subcategories over time.....	38
3-1 EarSketch Web Client Interface.....	49
3-2 EarSketch Web Client Interface (Modified).....	51
3-3 Example HMM Diagram.....	53
3-4 Hidden Markov Model Hidden States Density Plot (All Sessions)	57
4-1 Post-Survey Responses on Perceived Partner Contribution	68
4-2 HMM transition example.	69
4-3 Hidden Markov Model Hidden States Density Plot (Comparison)	72
5-1 EarSketch Interface	93
5-2 EarSketch Interface with CAI enabled	95
5-3 EarSketch CAI Architecture Diagram.....	95
5-4 CAI Chat Interface (Initial Prototype).....	99
5-5 Flow of information in the Wizard of Oz study.....	106
5-6 The CAI Interface Music Menu.....	109
6-1 Initial interface sketches for study conditions.....	118
6-2 Medium-fidelity Figma prototype.....	119

6-3	Final implemented interface design	120
6-4	Word Count Distribution by Condition	125
6-5	Semantic homogeneity methodological flowchart	135
6-6	Distribution of Semantic Similarity Scores by Condition	136
6-7	Attribution mapping of human-AI collaborative artifacts	157
6-8	Sequential attribution distribution in human-AI collaborative artifacts	158
6-9	Request Types by Attribution Quadrant	159
6-10	Post-Activity Likert Responses by Writing Condition	160
6-11	Participant Preference Between Writing Conditions	160
6-12	Participant Experience with AI Language Models	161
6-13	Text Length Comparison: AI-Assisted vs. Individual Writing	161
6-14	Semantic Similarity Dendrograms by Condition	162
B-1	Hidden Markov Model of the Co-Creative Process	172
D-1	Warmup Activity: AI Assisted Condition	174
D-2	Warmup Activity: Individual Condition	174
E-1	Writing Activity: AI Assisted Condition	175
E-2	Writing Activity: Individual Condition	175

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

UNDERSTANDING CO-CREATIVE PROCESSES THROUGH DIALOGUE ACROSS
HUMAN-HUMAN AND HUMAN-AI PARTNERSHIPS

By

Amanda E. Griffith

May 2025

Chair: Kristy Elizabeth Boyer

Major: Human-Centered Computing

The allure of artificial intelligence (AI) as a creative force, from Ada Lovelace’s musical visions to contemporary applications such as refining John Lennon’s vocals to create the last Beatles song, has long captured our imagination. Today, conversational AI, powered by large language models (LLMs), is capable of not just serving as tools but as genuine creative collaborators. However, we lack a clear understanding of how these technologies truly impact our creative processes and creative outcomes.

This dissertation investigates human-AI co-creativity through a series of interconnected studies spanning computational music creation and narrative writing. The research progresses from analyzing the co-creative process between human pairs in a musical context, to designing a co-creative AI agent based on these analyses, and finally examining the impact of contemporary conversational AI on narrative writing processes and products

The initial analysis from the human-pair studies, using hidden Markov modeling on collaborative dialogues and interaction logs, identifies seven distinct co-creative states and reveals how specific dialogue patterns correlate with partner satisfaction. These findings inform the design of a co-creative AI (CAI) assistant. The multi-year iterative design of CAI highlights several challenges in human-AI creative partnerships.

Building on these foundational studies in computational music creation and extending across creative domains, the research transitions to narrative writing—a more accessible creative

practice with distinct collaborative dynamics. In a culminating study, participants wrote stories with and without conversational AI assistance. Findings reveal that AI-assisted narratives exhibit significantly increased text length, lexical diversity, and lexical sophistication. However, these AI-assisted narratives also show significantly greater semantic similarity across texts compared to the individually-written narratives. Furthermore, an analysis of creative agency distribution reveals a striking polarization between users who maintain high control and those who delegate extensively to AI, with specific dialogue strategies strongly predicting these agency patterns. These results demonstrate that human-AI co-creativity is not uniform but encompasses diverse engagement patterns with distinct implications for creative processes and products.

The research concludes that effective integration of AI into creative practices requires careful consideration of interaction dynamics, user agency, and potential trade-offs. Key contributions include: (1) characterization of co-creative dialogue states in collaborative processes; (2) evidence linking dialogue patterns to both partner satisfaction and creative agency distribution; (3) design principles for co-creative AI systems; (4) insights into creative agency distribution in human-AI co-creativity; (5) analysis of AI's impact on creative artifacts; (6) development of research platforms for studying human-AI co-creativity; and (7) cross-domain insights on co-creativity principles. These contributions advance theoretical understanding and inform implementation of AI systems that support human creativity across multiple domains.

CHAPTER 1 INTRODUCTION

The expanding role of AI reflects a longstanding fascination with the potential of machines to perform not just logical computations but also to engage in creative processes—a vision that traces back to the 19th century computing pioneer, Ada Lovelace, who is credited as the first person to recognize that the “computing machine” had applications beyond pure number calculation [52]. She mused that computational machines would eventually be able to compose music [120]. This once-theoretical concept has now materialized in concrete applications, as demonstrated by AI’s capability to extract and refine John Lennon’s vocals for the production of a previously unfinished Beatles track [141], illustrating how far we have come in realizing these early musings on creative computing machines.

The evolution of technology’s role in creative tasks illustrates this trajectory. In music creation, we’ve seen progression from simple algorithmic composition tools [30] to AI systems capable of generating original compositions [121]. In writing, we’ve seen progression from basic tasks such as identifying typographical errors [134] to more sophisticated capabilities like automating text correction and sentence completion [13, 201] and even further, enabling text summarization [23] and predictive text responses for our emails [22, 81]. The rapid adoption of advanced conversational AI tools powered by large language models (LLMs), like ChatGPT [179], into our daily routines and tasks underscores the growing impact of AI on society and has shifted the paradigm from AI as a mere tool to a potential partner capable of creative collaboration between humans and AI. This shift holds significant implications not only for professional creative industries but also for educational contexts, where fostering creativity and collaborative skills are key objectives.

As we continue to integrate these advanced AI tools into our creative processes, fundamental questions arise about the nature of human-AI co-creativity: What dialogue states characterize the co-creative process? What is the relationship between the co-creative process and co-creative outcomes? How do we design AI agents to be co-creative partners? How does AI assistance influence creative outcomes and the creative process itself? This dissertation addresses

these questions through a series of interconnected studies spanning computational music creation and narrative writing, providing insights into human-human and human-AI co-creativity across different domains.

1.1 Research Motivation

This dissertation investigates co-creative processes and their outcomes through a series of studies in different creative domains. My research path began with investigating human-human co-creativity to establish foundational insights into co-creative processes, progressing to the design and evaluation of a co-creative AI partner, and culminating in an examination of how conversational AI influences narrative writing. The progression of the research is guided by five key research questions:¹

1. **Understanding Co-Creative Processes:** Establishing foundational knowledge about human-human co-creativity in computational music creation (Chapters 3 and 4)
 - **RQ1:** What are the dialogue states that characterize the co-creative process between learners, and how do learners transition between these dialogue states?
 - **RQ2:** What are the relationships between partner satisfaction and the dialogue patterns that comprise the co-creative process?
2. **Designing and Evaluating Co-Creative AI:** Applying these insights to inform the design of a co-creative AI partner for computational music creation and identifying key challenges in human-AI creative collaboration (Chapter 5)
 - **RQ3:** What challenges and opportunities emerge when designing and evaluating a co-creative AI for an expressive CS learning environment?
3. **Examining Impact on Creative Outcomes:** Extending insights from prior studies into a new domain, investigating how conversational AI assistance influences both the creative process and resulting artifacts in narrative writing (Chapter 6)

¹ The work described in Chapters 3-5 was supported by the National Science Foundation (NSF) under grants DRL-1813740 and DRL-1814083. I was supported by NSF & IES on grant 2229612 for the final two years of my dissertation work. Collaborators on this project include Gloria Ashiya Katuka, Joseph B. Wiggins, Amogh Mannekote, Erin Truesdell, Jason Smith, Sarah Mathew, Tom McKlin, Brian Magerko, Jason Freeman, and Kristy Elizabeth Boyer.

- **RQ4:** How does conversational AI assistance impact users' final narrative artifacts in terms of text length, linguistic complexity, and cohesion compared to narratives written individually?
- **RQ5:** How does conversational AI assistance influence users' creative decision-making and narrative development?

This research progression reflects an evolving understanding of co-creativity—from basic co-creative patterns to considerations of creative agency and the impact of AI assistance on creative artifacts, with the transition from studying co-creativity in the context of computational music remixing to narrative writing being motivated by several key considerations: (1) Narrative writing provides a more accessible domain for studying creative processes, allowing for broader participant recruitment while still offering rich opportunities to observe creative decision-making; (2) The field of AI-enhanced human creativity has seen a marked increase in focus on narrative writing tasks that coincides with recent advancements in LLMs, as highlighted in my literature review (Chapter 2); (3) Narrative writing, compared to other forms of writing, offers distinct advantages for studying creative processes due to its inherent openness to imagination and subjective nature, providing an ideal context (both for general creativity research and for understanding potential impacts in educational writing activities) for examining human-AI creative collaboration [178, 78].

The primary goal of this dissertation is to develop an understanding of human-AI co-creativity across different domains and contexts. By examining both computational music creation and narrative writing, this research identifies both domain-specific patterns and cross-domain principles of human-AI creative collaboration. This exploration is timely and pertinent as increasingly sophisticated AI tools are becoming ubiquitous, permeating various aspects of creativity and productivity, while there is still limited understanding of these tools' impact on human creative processes and outcomes.

In examining these creative capabilities of AI, it is essential to establish how creativity itself is conceptualized in this dissertation. I define creativity as the generation of ideas, artifacts, or solutions that are both novel (original, unexpected) and valuable (appropriate, useful) within their

context [171, 153]. In co-creative contexts, this definition extends to incorporate the collaborative development of creative outputs through interactions between multiple agents—whether human or artificial—where both partners contribute to the novelty and value of the resulting creation [84, 196]. This framing of creativity as both an individual and collaborative process underpins the exploration of human-AI partnerships throughout this work.

1.2 Contributions

This dissertation makes several contributions to the understanding of human-AI co-creativity:

1. **Characterization of Co-Creative Dialogue States:** Identification and analysis of seven distinct dialogue states that characterize the co-creative process between learners in computational music creation, providing a framework for understanding co-creative interactions
2. **Connection Between Dialogue Patterns and Collaboration Outcomes:** Evidence of how specific dialogue patterns and state transitions correlate with partner satisfaction in human-human collaboration, and how dialogue strategies (particularly the types of requests) predict creative agency distribution in human-AI narrative co-creation
3. **Design Principles for Co-Creative AI Systems:** Guidelines for designing AI systems that can effectively support both the technical and aesthetic aspects of creative work
4. **Understanding of Agency Distribution in Human-AI Collaboration:** Insights into how different users approach creative collaboration with AI, including patterns of agency maintenance and delegation
5. **Analysis of AI’s Impact on Creative Artifacts:** Empirical evidence of how AI assistance affects the linguistic characteristics and semantic diversity of narrative artifacts

These contributions advance our understanding of co-creativity, provide practical insights for the design and implementation of AI systems that support human creativity across multiple domains, and reveal the influence of this support of on creative artifacts.

1.3 Overview

The remainder of this dissertation is organized as follows: In Chapter 2, I present a systematic literature review of the current state of AI support for creative tasks, focusing on identifying the types of creative tasks AI tools currently support. Chapter 3 describes my published research that investigates the patterns of collaborative dialogue among high school student dyads engaged in co-creative tasks using a hidden Markov model, revealing seven co-creative states. Chapter 4 builds on Chapter 3 by examining how these co-creative states correlate with students' satisfaction with their partners. Chapter 5 documents the development of EarSketch CAI and presents design recommendations for co-creative AI systems. In Chapter 6, I present my dissertation research investigating the impact of conversational AI on narrative writing. Finally, in Chapter 7, I synthesize findings across studies, discuss their broader implications, and suggest directions for future research in human-AI co-creativity.

CHAPTER 2

LITERATURE REVIEW

The overarching question of my dissertation focuses on how AI can support human creativity. In this context, I'm particularly interested in how conversational AI assistance influences the creative writing process and the characteristics of the resulting narrative artifacts. This chapter reviews key research relevant to my dissertation, establishing the background and motivation for this review in Section 2.1 and detailing the review methodology in Section 2.2. Next, Section 2.3 discusses four primary creative task types creativity support systems support humans with: Design, Narrative, Idea Management, and Music/Audio tasks. Then, Section 2.4 discusses the research trends within these task categories, highlighting a notable surge in AI tools to support Narrative tasks in the recent years—an observation directly contributed to my decision to examine narrative writing in my dissertation work (Chapter 6). Finally, Section 2.5 provides an overview of methods for analyzing collaboration and the roles of intelligent systems that support collaboration, which is important for understanding the dynamics of human-AI co-creativity. Overall, this chapter provides a foundation for my dissertation work by identifying the types of creative tasks AI tools support and examining the relevant research trends and collaboration principles.

2.1 Background and Motivation

Building on decades of research in human creativity and computational support, this systematic review aims to understand the capabilities and limitations of AI in aiding human creativity, serving as a foundation for my dissertation. The primary research question for this review is: *What types of creative tasks do AI tools currently support?*

As AI technologies have advanced, computers have evolved from tools that merely support creative work to active participants in the creative process through co-creativity—a paradigm where humans and computers jointly contribute to creative artifact and idea production [35]. In this collaborative framework, both partners share responsibility for the creative output [58, 83], engaging in exchanges that shape the development of creative artifacts. However, the emergence of co-creative systems presents significant challenges in designing computational agents that can

effectively support users in open-ended, creative endeavors while maintaining meaningful collaboration.

The examination of AI-enabled creativity comes at a critical juncture in technological evolution, though this investigation faces notable challenges. The rapid pace of AI innovation often outpaces academic publishing, while advances in the private sector remain largely opaque. Nevertheless, this review identifies and analyzes four primary categories of AI-supported creative tasks—Design, Narrative, Idea Management, and Music/Audio—and traces their development over time. Through this analysis of current AI-enabled creativity and supported tasks, I aim to establish a foundation for understanding how AI transforms creative expression and to identify promising directions for future research.

2.2 Methods

This section documents the methodological approach of the systematic literature review, detailing the sampling strategy, screening process, and analytical framework used to examine AI-enhanced creativity support tools in HCI research from 2018 to 2023.

2.2.1 Sampling

This systematic literature review builds upon a 2019 review of creativity support tools in HCI [51], encompassing HCI research papers published between 2018-2023. The review protocol adheres to the guidelines of Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [122]. The flow diagram (Figure 2-1) documents each step of the process.

I began the identification phase with a database search in the **ACM Full-Text Collection**, expanding upon the original search terms from the 2019 review [51] by adding *Creative AI* and *Creative Agent*. The complete search string used was:

(Author Keyword: Creativity) OR (Any Field: Creativity Support Tool” OR Creative AI” OR “Creative Agent”)

The database search, completed on February 13, 2023, identified 483 unique records, all from peer-reviewed conference and journal publications. As shown in the PRISMA flow diagram (Figure 2-1), I performed an initial screening phase, during which I excluded 359 records for not

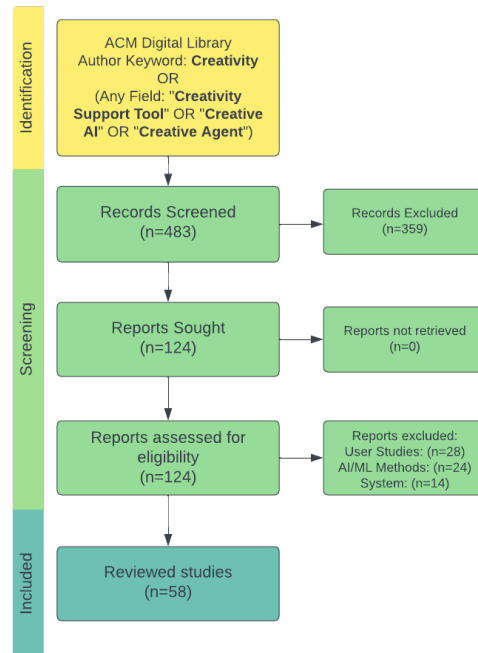


Figure 2-1. PRISMA Flow Diagram illustrating the literature screening, eligibility, and inclusion steps in the review.

meeting the basic inclusion criterion of introducing a new creativity support tool with AI or machine learning (ML) capabilities. When AI/ML applications were not explicitly stated, I maintained a conservative approach by retaining papers with plausible AI/ML implementations for further assessment.

The remaining 124 papers progressed to the eligibility assessment phase. I successfully retrieved all 124 papers for detailed review, as indicated by the “Reports not retrieved (n=0)” entry in the PRISMA diagram (Figure 2-1). I evaluated these papers against three specific exclusion criteria:

- **User Studies:** I excluded papers lacking user studies involving creative tasks with the tool. This exclusion criterion included papers without any user studies, studies where users did not interact with the tool, and evaluations focused solely on the support tool’s output (e.g., participants’ only roles in the study are rating how well a tool or model performs a creative task, rather than performing the creative task themselves.)
- **AI/ML Methods:** I excluded papers without clear descriptions of implemented AI/ML methods. This exclusion criterion included cases with ambiguous AI/ML usage or where AI implementation was proposed but not yet realized.

- **System:** I excluded papers based on tool characteristics, including those presenting minor modifications to existing tools or lacking adequate system descriptions. For systems documented in multiple papers, I retained the paper with the most comprehensive system description.

As documented in the PRISMA diagram (Figure 2-1), I excluded 66 papers based on the exclusion criteria: 28 due to the user studies criterion, 24 due to the AI/ML methods criterion, and 14 due to the system criterion. This systematic screening process yielded a final set of 58 papers for review.

2.2.2 Analysis

In this systematic literature review, I focus on the user tasks assisted by AI-enhanced creativity support tools. To identify and categorize these tasks, I anchor my analysis using an inductive thematic approach.

First, I thoroughly read each selected paper and extracted a concise description of the task the users completed during the user study. These tasks were often illustrated by the artifacts created, providing clear markers for my analysis.

After extracting the task descriptions, I recorded the descriptions to a virtual Miro board and grouped the descriptions based on the type of task performed. I firmly grounded my analysis on the users' active work, as opposed to potential tasks the tools could hypothetically facilitate. Some advantages to grouping the tasks based only on those performed in user studies include: (1) ensuring that the technology listed in the system descriptions genuinely supports the classified task types, (2) enhancing the objectivity of the analysis by eliminating any need for speculative conclusions regarding the authors' envisioned but untested use cases, and (3) assisting the maintenance of mutually exclusive task categories, enabling a simpler and more straightforward analysis.

Initially, fourteen categories emerged from this process. Afterward, I consolidated these fourteen categories into four main themes: Narrative, Design, Music, and Idea Management, shown in Table 2-1. I kept the 14 initial categories as subcategories of the four primary categories, shown in Table 2-2 with the papers.

Table 2-1. Literature Distribution Across Primary Categories: Distribution of reviewed papers examining AI-enhanced creativity support tools, organized into four thematic categories identified through inductive analysis (Design, Narrative, Idea Management, and Music/Audio)

Category	Papers
Design	[3, 46, 126, 65, 85, 161, 191, 160, 79, 67, 36, 132, 70, 105, 26, 90, 53, 158, 189, 4, 192, 107, 114, 74, 91]
Narrative	[33, 154, 28, 45, 71, 76, 113, 129, 170, 108, 56, 55, 198, 27, 96, 162, 24]
Idea Management	[57, 200, 68, 195, 98, 190, 97, 80, 172, 75, 99, 9]
Music/Audio	[156, 135, 139, 157]

Table 2-2. Literature Distribution Across Subcategories: Systematic categorization of research papers examining AI-enhanced creativity support tools, showing the distribution across fourteen distinct task domains and their consolidation into four primary categories.

Category	Subcategory	Papers
Design	Illustration	[3, 85, 191, 160, 132, 105, 53, 158, 192, 107, 91]
	3D	[65, 26, 90, 189]
	Game	[46, 67, 4]
	Multimedia	[126, 114]
	Digital	[161, 36]
	Fashion	[79]
	Interaction	[70]
	IoT	[74]
Narrative	Written	[33, 154, 28, 71, 170, 129, 113, 55, 56, 27, 162]
	Visual	[108, 76, 198, 96, 24]
	Spoken	[45]
Idea Management	Collage	[200, 98, 97, 75, 99, 9]
	Misc	[57, 68, 195, 190, 80, 172]
Music/Audio	-	[156, 135, 139, 157]

Section 2.3 describes each of the identified categories and their subcategories. The subcategory descriptions also feature brief descriptions of the included tools to provide context and a deeper understanding of the task types. Section 2.4 discusses temporal trends I observed in the Design and Narrative categories from 2018 to 2023 and potential influences on those trends.

2.3 Creative Task Support

This section elaborates on the four main categories identified in the systematic literature review: Design, Narrative, Idea Management, and Music/Audio. Each of these categories encompasses different types of user tasks assisted by AI-enhanced creativity support tools, which are further subdivided into subcategories. The Design category, which includes the largest

number of tools, involves tangible or observable products or experiences often with visual or interactive components. The Narrative category covers various forms of storytelling, showcasing how AI-enhanced creativity support tools can assist narrative creation across written, visual, and spoken mediums. Idea Management tools aim to enhance the creative ideation process, providing platforms for the generation, development, and management of ideas across various creative tasks. Lastly, the Music/Audio category highlights diverse applications of AI in sound and music creation, providing users with tools to manipulate and evaluate audio elements. Each of these categories and their respective subcategories are described in detail, presenting a broad perspective on how AI assists in a wide range of creative tasks.

2.3.1 Design

The Design category includes 25 tools and is the largest of the four categories. Tools in the Design category emphasize creating tangible or observable products or experiences, often involving visual or interactive components. These tools are AI-aided applications that assist users in the conception, development, and execution of visual, spatial, and interactive elements across various fields. Subcategories of Design, shown in Table 2-2, include illustration, game design, multimedia content creation, 3D modeling, digital design, fashion, interactions, and the Internet of Things (IoT).

2.3.1.1 Illustration

The Illustration subcategory is the most common task subcategory within the Design category encompassing 11 of the 25 Design tools. The Illustration subcategory focuses on AI tools specifically supporting illustrative tasks, such as drawing and sketching. These tools leverage AI and machine learning techniques to assist users by providing inspiration [85, 105], assisting in learning drawing skills [191, 192, 107], increasing the accessibility and usability of design processes [91], or facilitating sketching of specific elements, such as emotional expressions [160]. Others utilize AI to provide unique drawing experiences, such as enabling users to paint with physical objects [53] or applying textures from live video footage to their artworks in real-time [158]. Additionally, some tools in this category enable a unique form of co-creation, in

which humans and AI collaborate directly on the same artifact to create drawings together [3, 132]. Tools from the Illustration subcategory showcase the transformative potential of AI in redefining the landscape of artistic expression.

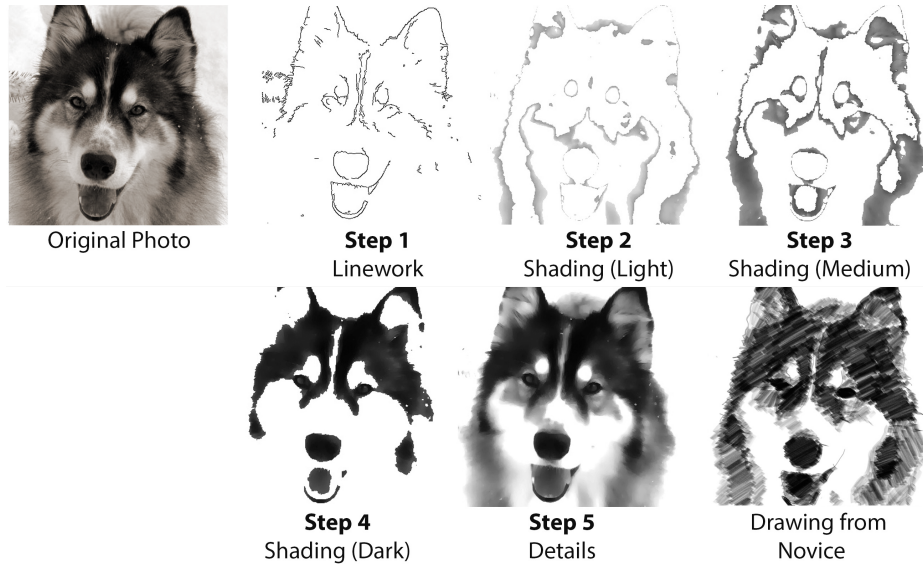


Figure 2-2. Depiction of the interactive drawing tutorial generated by the DrawMyPhoto system [191].

2.3.1.2 3D

The 3D Design subcategory consists of four Design tools that focus on aiding users in creating, modifying, and visualizing three-dimensional objects and spaces. These tools employ AI techniques to help users, especially novices, in various tasks, such as providing users with adaptive AI-generated suggestions for interior colorization [65], creating 3D models using body movements [26], exploring and combining design alternatives for garden design [90] (shown in Figure 2-3), and personalizing everyday objects with custom circuit layouts [189]. This variety of tools underlines the diversity of applications within 3D Design, demonstrating how AI can assist users in designing and manipulating 3D spaces and objects with increased ease and creativity.

2.3.1.3 Game

The Game Design subcategory consists of three distinct tools: Morai Maker, Questgram, and Cine-AI, each demonstrating unique applications of AI in augmenting and simplifying different aspects of game creation, including level design, quest design, and cutscene generation.



Figure 2-3. Garden Design Interface (top) and Garden Exploration Interface (bottom) of Mixplorer [90].

Morai Maker, shown in Figure 2-4, is designed for game designers, both professional and novice, and offers a platform for the design of Super Mario Bros.-style games with an AI partner [67]. Questgram is a mixed-initiative quest generation system intended for game developers and designers [4]. Its AI capabilities facilitate the creation of intricate quest structures for dungeon-style games, effectively streamlining and enhancing the complexity and depth of quest design. Cine-AI aims to serve independent game developers by procedurally generating camera behavior for in-game cutscenes [46]. The tool uses AI to automate the traditionally complex and

time-consuming task of cinematography in game development, enabling developers to focus on other aspects of their games.

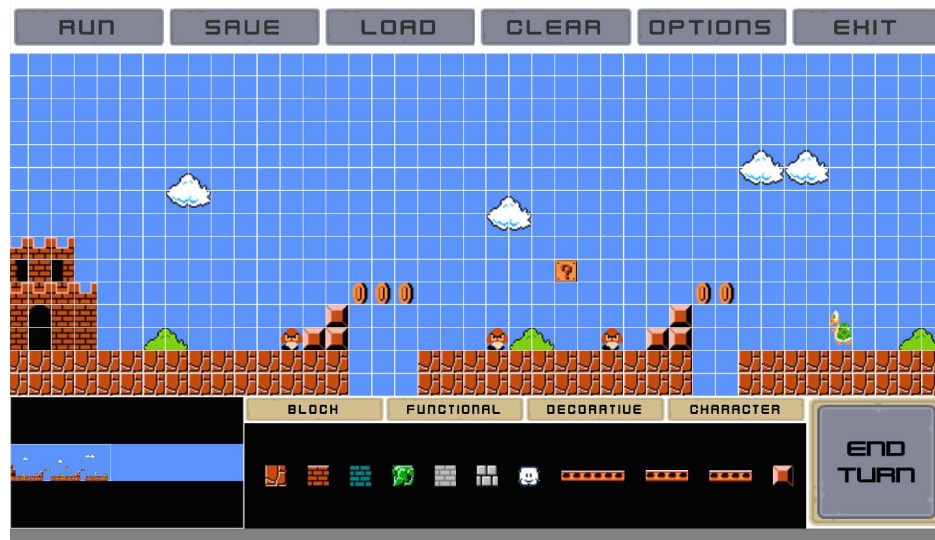


Figure 2-4. Screenshot of the Morai Maker level editor [67].

These tools cater to diverse game design elements and underline the breadth of AI's capabilities in game development. They demonstrate how AI can automate the production of cutscenes, aid in the design of game levels, and streamline quest creation. Collectively, they exemplify the vast array of tasks that AI can support within this field, thereby enhancing both the creativity and efficiency of game design.

2.3.1.4 Multimedia

The Multimedia subcategory consists of two tools that focus on creating and curating multimedia content with the assistance of AI. CreaTable, shown in Figure 2-5, designed for people with aphasia, is a tangible platform for creating and curating multimedia content providing an inclusive platform for creativity and content production [126]. PictureBlocks, on the other hand, is intended for young children who are still developing their literacy skills [114]. It transforms text inputs into the corresponding sprite objects, allowing children to create digital pictures by typing words. This provides a creative and engaging way for children to learn and improve their literacy skills. Collectively, these tools underscore the versatility of AI in aiding the creation and curation of multimedia content, catering to a range of user groups and contexts. They demonstrate how AI

can assist in overcoming various challenges, from making multimedia content creation more accessible for people with aphasia to enhancing literacy development in children through interactive, multimedia-based learning experiences.

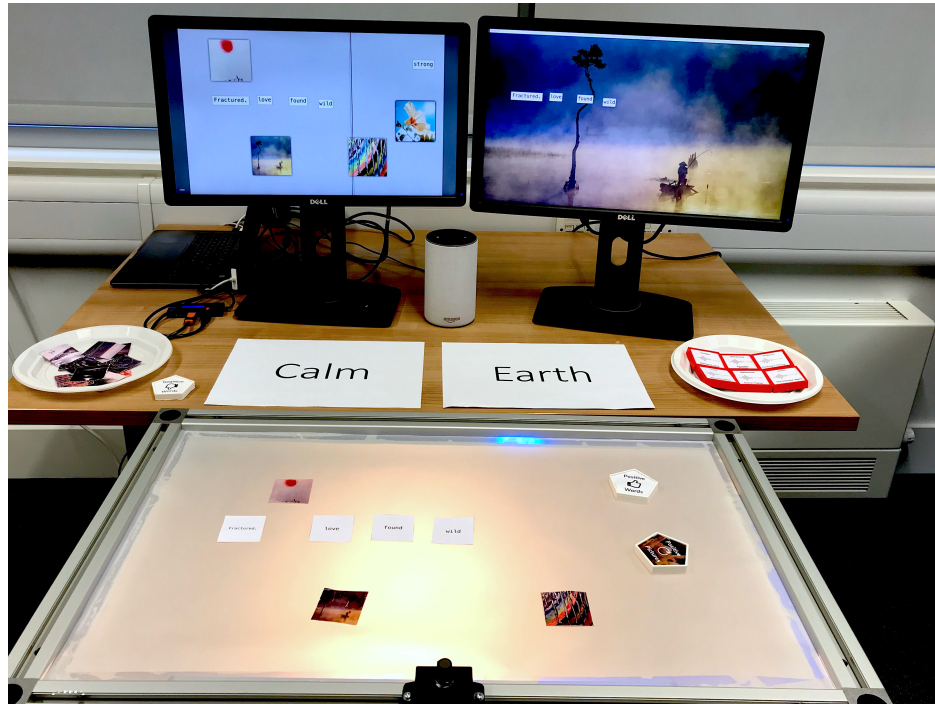


Figure 2-5. The CreaTable media sequencer system [126].

2.3.1.5 Digital, Fashion, Interaction, and IoT

The remaining Design tools fall into smaller yet unique subcategories: Digital Design, Fashion, Interaction, and the Internet of Things (IoT). The Digital Design tools use AI to guide exploration in high-dimensional design spaces aiding users in design tool creation [161], and to generate diverse layouts for graphical user interfaces (GUIs) [36]. In Fashion, AI assists users in overcoming ideation obstacles during design [79]. The Interaction category boasts a tool for designing immersive movement interactions [70], while IoT leverages AI to visualize everyday objects' social networks [74]. Despite their fewer numbers, these tools show the growing intersection of AI with various design domains, offering innovative and creative opportunities.

2.3.2 Narrative

The Narrative category includes 16 tools encompassing various forms of storytelling. AI tools within this category aim to support and enhance narrative creation and consumption across different mediums. The use of artificial intelligence in this category leverages machine learning, computer vision, and natural language processing (NLP) techniques to support users in their creative process, providing them with inspiration, structural support, stylistic guidance, character creation assistance, and critique. The Narrative Task category spans a wide range of AI tools designed to aid users in their creative endeavors across written, visual, and spoken mediums.

2.3.2.1 Written

The Written subcategory remains the most common form of narrative task, with 11 of the 16 tools assisting in written narrative tasks. These tools employ NLP and machine learning techniques to aid writers in creating, refining, and exploring narratives in written form, such as facilitating planning, structuring, and generating and developing ideas [33, 154, 28]. Some tools provide stylistic and metaphorical assistance, enhancing the creative aspect of the narrative [170, 55]. Others offer unique insights by enabling writers to evaluate the representation of characters in their stories through visual analytics [71]. A few tools further enhance the narrative process through interactive guidance, unveiling novel storytelling perspectives, and cultivating rich, compelling sentences [129, 113, 56]. Other tools offer a combination of ideation and writing suggestions [27, 162]. Overall, these tools provide a range of support, from brainstorming to the execution of ideas, to enrich the user's written narrative experience.

2.3.2.2 Visual

While less common than its Written counterpart, the Visual subcategory is a burgeoning area in human-AI co-creative narrative tasks. It comprises four of the 16 tools in this category, notably accounting for a third of the tools introduced in 2022. These tools enrich visual narratives by facilitating the ideation, creation, and refinement of visual content. They assist users in finding relevant visual concepts and generating illustrations based on an article's tone, keywords, and aesthetic [108], collaborate with children in visual storytelling [198], and assist webtoon writers

and artists in their collaborative revision process by generating a variety of reference images, shown in Figure 2-6 [96]. Adding to this repertoire is an innovative agent that participates in an improv game with humans in the VR installation Robot Improv Circus, dynamically interpreting and executing actions to contribute to the visual narrative [76]. Doc2Video [24] improves the speed at which creators can make tutorial videos from programming cookbooks by creating a video prototype from a document. The tools in this subcategory thus contribute significantly to visual narratives, aiding in ideation, creation, and refinement processes.

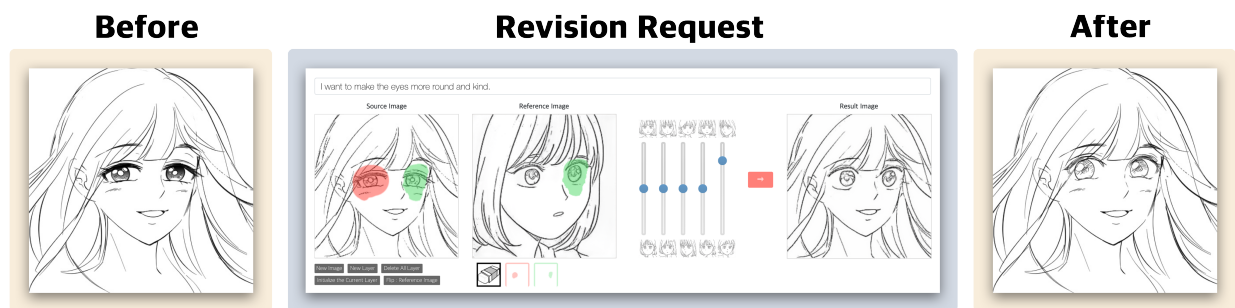


Figure 2-6. The We-toon system, a tool developed for NAVER WEBTOON, for revisions: artists sketch and send to writers (Before); writers modify the image to provide feedback (Revision Request); artists make changes (After). [96]

2.3.2.3 Spoken

Although smaller with just one tool, the Spoken subcategory focuses on the transient nature of oral storytelling, emphasizing the immersive and interactive experience over creating material or lasting artifacts. The tool represented in this subcategory, CreativeBot, featured in Figure 2-7, is an autonomous robot that engages children in collaborative storytelling activities [45]. Despite the limited size of this subcategory, it represents a unique and exciting frontier in narrative co-creation. The potential of AI to enhance real-time spoken storytelling points to a future where AI can contribute to more spontaneous and playful forms of co-creation.

2.3.3 Idea Management

The Idea Management category encompasses various tools to enhance the creative ideation process. These tools stimulate and manage idea generation and provide a means for organizing and visualizing these concepts. However, unlike tools in other categories that aid in ideation or

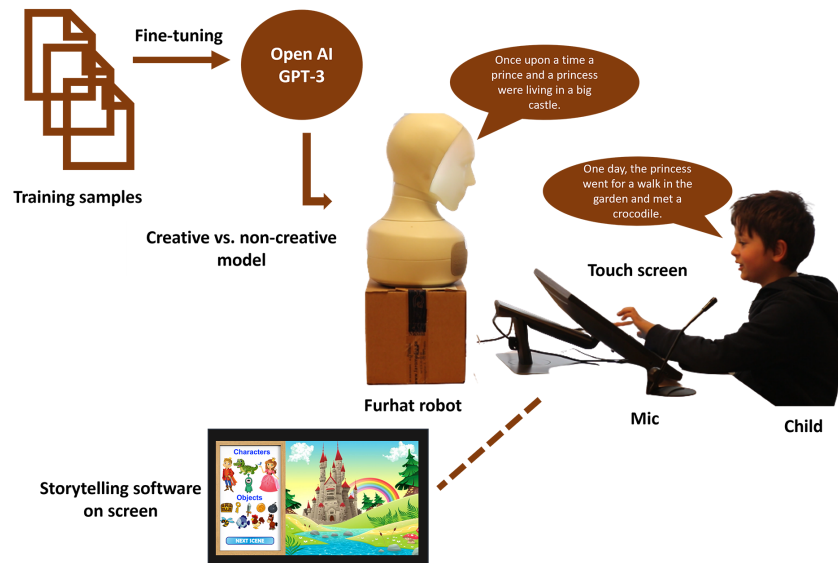


Figure 2-7. Interaction flow between the child and robot during the storytelling activity [45].

visualization as a means to facilitate specific tasks, tools in this category are designed to support the general process of idea generation, development, and management across a range of creative tasks. They do not restrict the user to a specific end product but instead provide a platform to cultivate, organize, and communicate ideas in diverse contexts.

The Collage subcategory includes six tools that support the process of collecting, organizing, and visualizing ideas. These tools assist users in creating collages or mood boards, integrating visual materials to encapsulate a mood, theme, or concept. They utilize AI and machine learning to offer the capability to repurpose image components [200] and manipulate multimedia in a 3D environment [75]. Still, most of the tools employ various methods to suggest relevant content [97, 98, 99, 9]. They emphasize the value of diversity and serendipity in the creative process, aiding users in making unexpected connections and gaining novel insights.

The remaining tools in this category have unique functionalities to support idea management. Some of these tools facilitate brainstorming and ideation by leveraging user-generated content to inform user models and provide adaptive suggestions [57, 195, 172]. Other tools in this category enrich the ideation process by providing insightful feedback on ideations using explainable AI [190] and tracing the 'thinking path' for visual metaphor ideation [80]. One tool delves into unconventional avenues for idea generation by manipulating dream

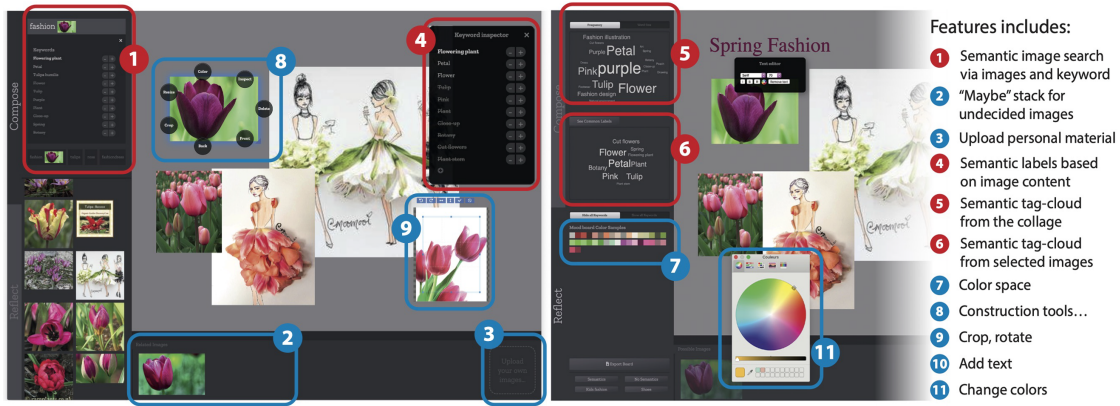


Figure 2-8. SemanticCollage: a semantically enriched digital mood board tool for image collection and interpretation. Blue tools support image and text manipulation; Red tools provide semantic labels [99].

content to spark creativity [68]. These tools showcase the diverse ways in which AI can assist users in managing their creative ideas, promoting innovation, and fostering creativity.

2.3.4 Music/Audio

The Music and Audio category highlights the diverse applications of AI in enhancing various aspects of sound and music creation. These tools assist users in exploring sound parameter spaces for tailored sound design [156] (Figure 2-9), providing an interactive platform for musicians to rehearse and perform in a virtual reality environment [135], generating line-level pop lyrics to aid songwriters in the creative process [139], and offering real-time visual feedback on audio quality during voice recording [157]. By integrating AI technologies into these tools, creators can utilize, manipulate, and evaluate audio elements, thus expanding the possibilities and potential for innovation in audio and music domains.



Figure 2-9. The CoExploer interface featuring standard parametric knobs for direct manipulation and real-time agent action visualization, an interactive history for state revisiting and feedback tracking keyboard inputs for feedback communication and agent exploration control, and a reset button to clear agent memory [156]

2.4 Trends in Task Support

This section examines the temporal evolution of AI-enhanced creativity support tools, focusing particularly on the shifting patterns in Design and Narrative tasks from 2018 to 2023. First, I present an overview of the general trends observed across all categories. Then, I analyze in detail the progression of Design tools, followed by an examination of the Narrative category’s development. Finally, I discuss the implications of these trends and potential future directions for research in this domain.

This analysis reveals significant trends in research focus and technological advancement, providing insights into the future direction of AI-supported creativity. As Figure 2-10 shows, the Design category saw a significant surge in 2020, with the number of papers more than doubling from the previous year, demonstrating a rapid growth in interest and research output within this category. However, the following years show a decline in the number of Design papers. This surge and subsequent decline in Design category papers prompts interesting questions about the changing landscape of this research area and the factors influencing these shifts. In section 2.4.1, I will further explore this trend and discuss potential influences such as the increased diversification of tasks and the COVID-19 pandemic’s impact.

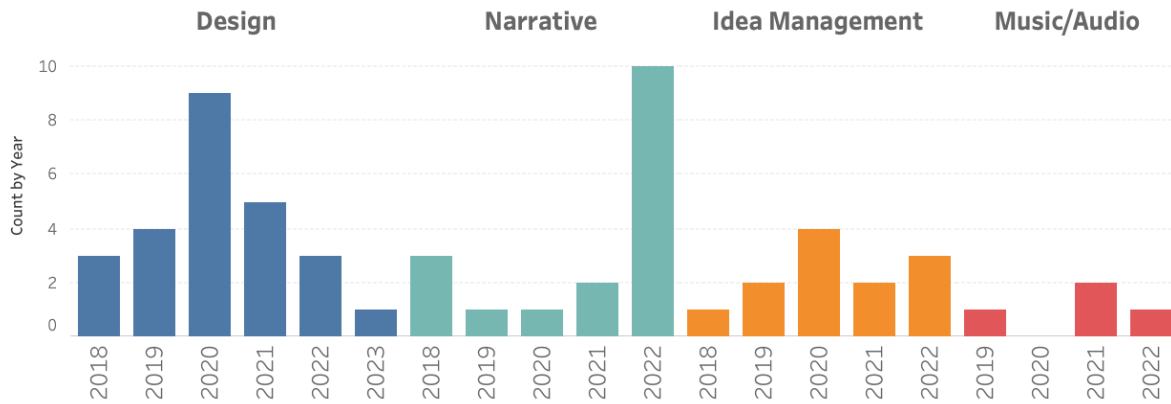


Figure 2-10. Temporal Trends in Human-AI Co-Creation Research. This figure displays the number of research papers published in four key categories—Design, Narrative, Music, and Idea Management—by year from 2018 to early 2023.

Conversely, the Narrative category demonstrates a contrasting pattern. Starting on par with the Design category in 2018, the interest in Narrative research seems to ebb for a couple of years, with only one paper published in both 2019 and 2020. However, the following years show a resurgence, particularly in 2022, which witnesses explosive growth in the number of papers related to Narrative tasks, a more than four-fold increase from the previous year. This recent surge of papers in the Narrative category raises interesting questions regarding the factors behind its growth. In section 2.4.2 I will explore how recent advances in language modeling and increased access to these models may have influenced these trends.

2.4.1 Design Over Time

As illustrated in Figure 2-11, the Design category has transformed over the last 5 years. Design-related research peaks in 2020, dominated by Illustration tasks. However, despite a subsequent decline in the number of Design papers, the category diversifies across subdomains.

The initial focus on understanding human-AI interaction and accomplishing basic tasks, primarily in the Illustration subcategory, demonstrates the nascent stage of AI assistance in creative tasks, providing a solid foundation for future advancements. For example, DuetDraw [132] examines the user experience of AI-embedded interfaces, enabling humans and AI to collaborate directly on the same artifact to create drawings together. As time progresses, the tools

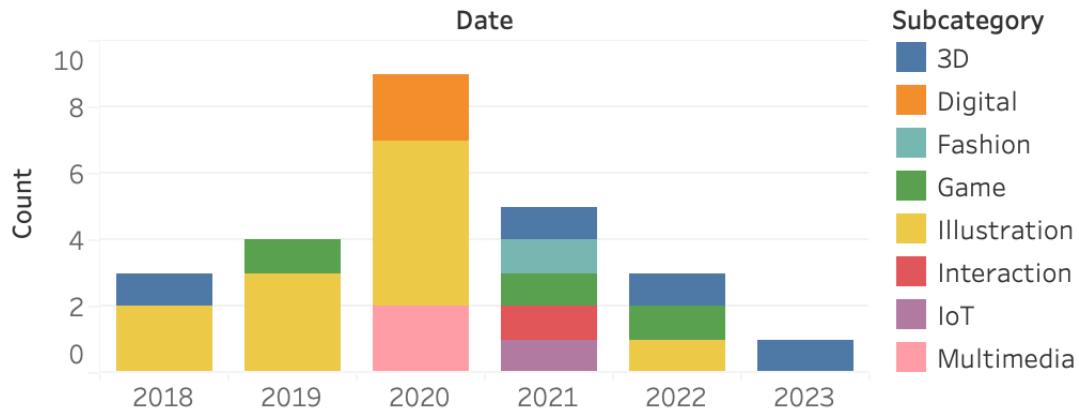


Figure 2-11. Distribution of research focus in Design subcategories over time. This figure illustrates the change in research attention on various design tasks from 2018 to 2023.

shift towards more complex and diverse subcategories such as 3D design, game design, and multimedia, signifying a maturation in the field, and possibly driven by advances in machine learning techniques and their application to specialized creative tasks. For example, IntuModels [26] helps novices create 3D models using their body movements leveraging CV, affinity propagation, dimensionality reduction, and motion sensing.

The shift from Illustration tasks and the decline in Design tasks overall may also suggest the natural progression of AI research in Design. Illustration tasks could have served as an intuitive starting point for understanding human-AI interaction in design tasks.

The COVID-19 pandemic may have influenced research trends in the Design category, particularly due to the specialized hardware requirements of many design tools that complicated remote access. Analysis of the 25 Design studies revealed that only six incorporated remote participation options. Remote-enabled studies maintained a consistent distribution across the examined time frame: two studies in 2019, one in 2020, one in 2021, and two in 2022. The stable, but limited, number of remote studies following the onset of the pandemic in 2020 may indicate systemic barriers to adapting design tools for remote research. This pattern could reflect underlying challenges in conducting design research remotely, as specialized hardware requirements and the need for controlled experimental conditions often favor in-person participation.

Remote participation options primarily appear in studies examining tools designed for professional users, where the tools function as extensions of existing software platforms. Cine-AI exemplifies this pattern: developed as a Unity plug-in for indie game creators, the tool's integration with established development environments allowed users to download and operate the software on their own computers within their standard workflow [46]. This integration with familiar professional tools may have reduced barriers to remote study participation by minimizing specialized equipment requirements.

The methodology for mapping human-AI co-creative tools established specific inclusion criteria requiring direct user engagement with creative tasks, distinguishing tools designed to augment human creativity from those focused solely on task automation. This methodological requirement may affect the observed patterns and trends, particularly regarding studies that do not adapt their protocols for remote participation.

2.4.2 Narrative Over Time

The analysis of Narrative tools, as illustrated in Figure 2-10, reveals a significant shift in research focus over the examined period. Publication patterns from 2018 through 2021 show relatively consistent output levels—three, one, one, and two papers respectively—followed by an increase to ten papers in 2022. Figure 2-12 indicates this increase coincides with the emergence of Visual and Spoken Narrative subcategories. However, while these new subcategories contribute to the overall increase, the Written Narrative subcategory also experiences substantial growth in 2022, surpassing publication counts from each preceding year in this review. This broader expansion across subcategories may relate to two concurrent developments in the field: advances in AI and machine learning capabilities, and increased accessibility to these advanced tools.

2.4.2.1 Advancement of AI tools

Recent advancements in AI and machine learning, specifically in the field of Natural Language Processing (NLP), have likely contributed to the observed trend of increased scholarly interest in Narrative tasks and allowed for more complex interactions. This is clear from both the technology and task progression.

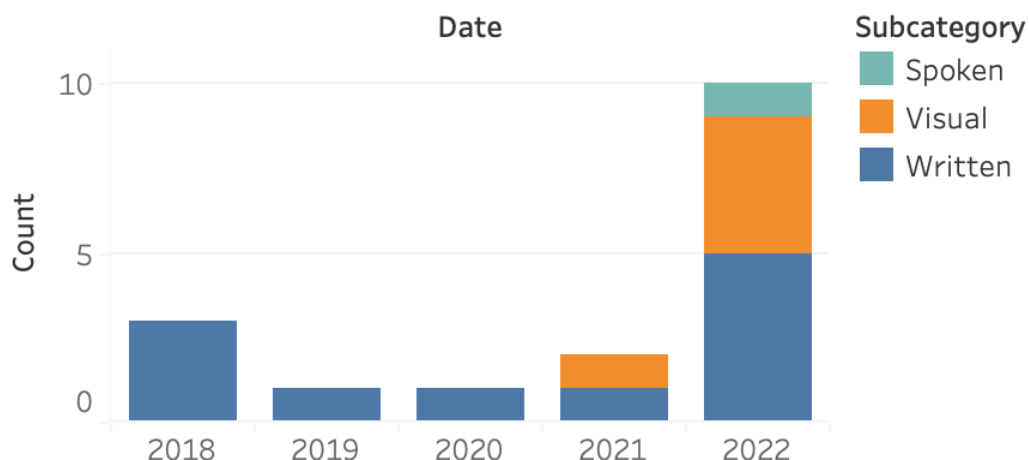


Figure 2-12. Distribution of research focus in Narrative subcategories over time. This figure illustrates the change in research attention on various narrative tasks from 2018 to 2023.

The analyzed tools show an evolution in AI/ML approaches over time. Tools before 2021 primarily use RNN or CNN architectures, word embeddings, and traditional NLP methods like named entity recognition and dependency parsing. For example, [28] developed a sentence suggestion system that combined an RNN-based language model with syntactic patterns from Wikiquote. Similarly, Metaphoria [55] used word embeddings and knowledge graphs to generate metaphorical connections based on semantic similarity. These early tools generated suggestions within constrained syntactic patterns.

In 2021, several tools incorporated transformer-based architectures. CharacterChat [154] implements TransferTransfo [194], a GPT-1 based model, for open-ended character exploration while using rule-based prompts and knowledge graphs for character definition. This combination of structured and open-ended interactions represented a new approach to co-creative tools.

Analysis of tools published in 2022 reveals a broader adoption of transformer architectures. Six of the ten Narrative tools published that year implement transformer models: four in Written Narrative, one in Visual Narrative, and one in Spoken Narrative. The Written Narrative tools utilize various transformer architectures: T5 [33], GPT-2 [56, 162], and GPT-Neo [27]. Additionally, two tools use GPT-3 [17]—one in Visual Narrative and one in Spoken Narrative—demonstrating applications beyond text generation.

The implementation of GPT-3 varied between tools, as illustrated by comparing Opal and CreativeBot.

Opal [108], designed for professional news illustrators, employs GPT-3 in tandem with Sentence-BERT and VQGAN+CLIP to support visual content generation based on an article's tone, keywords, and associated artistic styles. The system processes article text through GPT-3 to generate suggestions for keywords, tone, and artistic style. Users select from these suggestions to inform image generation parameters. Then, Sentence-BERT performs asymmetric semantic search to generate related terms, producing three variations of prompts that VQGAN+CLIP then uses to generate images.

On the other hand, CreativeBot [45] implements GPT-3 for interactive storytelling with children through an embodied robot interface, Furhat. The system uses a fine-tuned GPT-3 model to generate story continuations based on children's input. During turn-based storytelling sessions, GPT-3 processes the child's contributions by filtering irrelevant words, identifying thematic keywords, and generating contextually appropriate story segments. The Furhat interface presents these contributions while displaying programmed expressive behaviors, including head movements, eye gaze, and facial expressions.

These two implementations demonstrate distinct approaches to human-AI co-creation. Opal employs GPT-3 primarily for initial ideation and parameter selection, while CreativeBot utilizes it for continuous interactive narrative generation. The tools also serve different contexts and user groups: professional illustrators versus children engaged in storytelling. This diversity in implementation suggests the adaptability of large language models to various creative tasks and user needs.

These recent narrative systems differ from earlier narrative systems in their direct contribution to creative artifacts rather than solely providing suggestions. However, further research is needed to understand the implications of this shift toward more active AI participation in the creative process.

2.4.2.2 Rise of AI platforms

The increased adoption of transformer models correlates with improved accessibility through platforms like Hugging Face’s transformers library [193]. Hugging Face provides an open-source ecosystem with over 30,000 pre-trained models and MLOps solutions for AI deployment. Analysis of the reviewed papers reveals widespread utilization of this platform in recent Narrative AI research.

Six of the reviewed papers incorporate Hugging Face tools: five papers [33, 56, 154, 27, 162] retrieve language models directly from the platform, while one [71] implements NeuralCoref through spaCy’s NLP pipeline. Among these, four papers utilize Hugging Face models as alternatives to GPT-3. The remaining two papers implement TransferTransfo [194], based on GPT-1 [154] and T5 [138].

Analysis of publication timelines suggests a relationship between model availability and research implementation. Papers utilizing GPT-3 appeared in late 2022 [45, 108], following its unrestricted release in November 2021. Similarly, research implementing GPT-Neo appeared in April 2022 [27], after its March 2021 release, while papers using GPT-2 were published in early to mid-2022 [162, 56], following its full release in late 2019. These patterns suggest a relationship between model availability and research tool development, considering the natural time lag between technology availability and its appearance in published research, due to the time necessary for building, testing, and publishing research.

2.4.3 Implications and Future Directions

The insights from this systematic literature review suggest several implications for researchers and practitioners working with human-AI co-creative tools. As AI-enhanced tools expand across creative domains, research opportunities emerge to examine their long-term influence on creativity, learning, well-being, and cultural expression. Additionally, the field would benefit from comprehensive human-centered design guidelines that integrate insights across research areas. Understanding the variations in impacts and design requirements across creative categories remains an important area for investigation.

The evolution of design tools reveals efforts to support diverse user groups, including novices, professionals, young children, and people with disabilities, and highlights the potential of AI applications to foster inclusivity and accessibility in creative practices. However, there is a notable gap in addressing accessibility needs—among the reviewed tools, only CreaTable [126] specifically addresses disability accommodations. Given that approximately one in four adults in the United States live with some form of disability [20], this underrepresentation suggests an opportunity for HCI research to contribute to more inclusive design approaches.

The relationship between AI model accessibility and tool development, particularly in Narrative tasks, highlights the potential role of AI democratization in fostering innovation. This observation raises questions about creative support in resource-constrained settings. Research investigating how to enhance creative support for users with limited access to high-end hardware or high-speed internet connections could expand the reach of AI benefits.

This systematic literature review provides insight into the rapidly evolving landscape of human-AI co-creative tools. The analysis reveals increasingly sophisticated AI capabilities, from basic task assistance to complex creative functions. These findings inform my upcoming doctoral dissertation research examining how AI influences creative expression. As AI tools become more integrated into daily activities, understanding their impact becomes increasingly important for guiding responsible development and implementation.

While the landscape of AI-enabled creativity presents complexities, this review identifies key trends in Design, Narrative, Idea Management, and Music/Audio tasks. The observed patterns in Design and Narrative categories over five years suggest evolving research priorities. The Design category demonstrated initial growth followed by diversification into specialized subcategories, while the Narrative category showed recent expansion coinciding with advances in language models and increased accessibility to AI tools.

These trends raise important questions about the integration of AI into creative practices. As AI transitions from an assistive tool to a potential creative partner, integrating more deeply into creative workflows, understanding the nuances of this relationship becomes increasingly

important. Future research should examine not only technical capabilities, but also broader implications for creative expression, learning processes, and collaborative practices. This expanded scope of investigation reflects the growing complexity of human-AI creative partnerships and their potential impact on various domains of practice.

Additionally, this evolution in creative technologies occurs against a backdrop of rapid innovation that often outpaces academic publishing and remains partially obscured by private sector developments. However, these challenges underscore rather than diminish the importance of systematic investigation into AI's role in creative processes. Such research can help shift discourse from concerns about technological displacement toward evidence-based examination of optimal implementation. This reframing presents opportunities for HCI research to guide the integration of emerging technologies while maintaining focus on human-centric values and ethical considerations.

This systematic examination of AI-assisted creativity provides a foundation for investigating how these technologies influence creative expression and shape human-AI creative partnerships. As the field continues to evolve, maintaining focus on empirical investigation while considering ethical implications will be crucial for developing systems that genuinely enhance human creative capacity.

2.5 Collaboration

While previous sections examined the types of creative tasks AI tools support and their evolution over time, this section explores the fundamental principles of collaboration that inform the design of AI creative partners. First, I examine frameworks for analyzing collaboration, drawing from established research in human-human collaboration. Then, I discuss how these principles can be applied to developing effective collaborative AI systems. Understanding these principles is crucial for developing systems that can effectively engage in co-creative processes with humans.

2.5.1 Collaboration Analysis

Drawing from established research in human-human collaboration, this section examines frameworks for understanding collaborative processes that can inform AI system design. For many years researchers have attempted to characterize collaboration through identifying group processes and interaction patterns. Work from [10] provides an early example of investigating how group processes during problem-solving were distributed over time, finding that groups move through phases of collaboration and that different group processes characterize these phases. Since then, researchers have analyzed many types of collaborative interactions using a variety of coding schemes. Research on interaction processes generally uses sequences of low-level codes (e.g., dialogue acts) to identify patterns and higher-level constructs, such as cycles of turn taking [144], ideal communication cycles [184], knowledge sharing events [169], inquiry threads [199] and collaborative episodes [25]. Identifying patterns in group processes allows researchers to distinguish differences between effective and ineffective groups. [184] found that high-performing groups had a higher proportion of ideal communication cycles (i.e., beginning with a discussion about task preparation and ending with an evaluation) than low-performing groups. [168] identified collaborative learning behaviors that were exhibited by effective collaborative learning teams. [25] explored how individual and collaborative patterns were associated with individual outcomes such as frustration, technical proficiency, and time spent in the makerspace.

2.5.2 Intelligent System Support

Building on these insights into human collaboration, researchers have developed various approaches for intelligent systems to support creative partnerships. Applications of co-creative agents range from improving collaborative skills, such as encouraging “deep thinking” and initiative taking [72], to providing emotional support for elementary school students learning to code [124]. A co-creative agent can support human creativity through supporting the creation of music through code [180] and supporting collaborative ideation through sketches [106]. For agents acting as co-creative partners, students can benefit from the knowledge scaffolding provided by collaboration described by [94] by having a partner that can provide the right

knowledge at the appropriate time. For example, providing relevant worked or partial worked examples [202, 117] and encouraging good coding and debugging habits as they move through the project are ways that can reduce the cognitive load of the technical portion of the task and allow students to spend more time on creating. To help with the aesthetic elements, co-creative agents could use details from the artifact and make suggestions or give examples for moving what the student already has done towards a completed version of the artifact. For example, if students are just inserting sounds without progressing to a song-like structure, agents could suggest a custom function to make a chorus with sounds already present in the code editor or even provide one. This could give students a clear direction for moving forward on completing their artifact.

Collaborative support systems are systems where the primary role is to support collaboration between humans. Some of these systems perform functions such as encouraging productive dialogue during collaborative learning [43, 177], while others help students regulate their collaboration by prompting students to use different reflection strategies [150]. Collaboration support systems can identify when pairs encounter issues through dialogue features [60], such as dialogue that decreases when errors increase, and encourage partners to communicate. They can even help guide the conversation when pairs are not having a productive conversation [43]. These systems can also encourage pairs to set explicit goals or tasks—like using a custom function—and help them track which ones they’ve accomplished. Similar to the co-creative agent, these systems could also suggest more advanced code structures if the pair seems not to be making progress towards completing the artifact.

2.6 Key Takeaways and Research Trajectory

This chapter provides a systematic overview of the landscape of AI tools designed to support human creativity, identifying key task domains (Design, Narrative, Idea Management, Music/Audio) and analyzing research trends within these areas. The review highlights a significant recent surge in AI tools focused on Narrative tasks (Section 2.4.2), coinciding with advancements in large language models. This trend underscores the timeliness and relevance of

investigating AI's impact specifically within the domain of narrative writing, a key focus of this dissertation's later chapters.

Furthermore, the review examines foundational concepts in collaboration analysis (Section 2.5) and the principles guiding intelligent system support for collaboration (Section 2.5.2). Understanding these dynamics is important as the effectiveness of human-AI co-creativity hinges not just on the AI's capabilities but also on the quality of the interaction and the structure of the partnership.

The insights from this literature review directly motivate and inform the empirical studies presented in the subsequent chapters. The identified gaps and trends shape the research questions and methodological choices that drive this dissertation's investigation into human-AI co-creativity: Chapters 3 and 4 into the fundamental nature of co-creative dialogue, addressing the need identified in the literature for a deeper understanding of collaborative processes (Section 2.5). By analyzing human-human collaboration in computational music creation, these chapters establish a baseline understanding of co-creative states and their relationship to collaborative outcomes (partner satisfaction), providing empirical grounding for designing and evaluating AI partners. Next, Chapter 5 applies these foundational insights to the practical challenge of designing a co-creative AI agent (CAI) for the EarSketch environment. This chapter directly engages with the principles of intelligent system support discussed in Section 2.5.2, documenting the iterative design process and evaluating the challenges and opportunities that arise when implementing AI support in an open-ended, creative learning context. Finally, Chapter 6 transitions the investigation to the domain of narrative writing, directly responding to the research trends identified in Section 2.4.2. Building on the cross-domain insights from the preceding chapters, this culminating study examines the impact of contemporary conversational AI on narrative writing processes and products, focusing on aspects like linguistic features, semantic homogeneity, and the distribution of creative agency.

By progressing from foundational studies of human collaboration to the design of AI partners and finally to an investigation of AI's impact in a widely accessible creative domain, this

dissertation aims to build an understanding of human-AI co-creativity, contributing to the theoretical knowledge and practical design principles outlined as necessary future directions in Section [2.4.3](#).

CHAPTER 3

DISCOVERING CO-CREATIVE DIALOGUE STATES DURING COLLABORATIVE LEARNING

This chapter presents a first step toward understanding how conversational AI can augment or support human creativity. The data from this study was collected and analyzed as part of an effort¹ to create a co-creative AI partner. For this study, I use the sequences of textual dialogues and system interactions between high school pairs to learn a hidden Markov Model. From the model, we can identify the co-creative states that characterize the co-creative process and how the student pairs moved through them as they collaboratively remix music in EarSketch.² Understanding the co-creative states and the transitions between them not only provides insight into the co-creative process between humans, it can also inform the design of co-creative AI tools can facilitate and actively contribute to such co-creative endeavors.

3.1 Background and Motivation

There is growing interest in using artificial intelligence (AI) to support collaborative learning. AI companions have the potential to improve learners' collaborative skills by, for example, encouraging “deep thinking” and initiative taking [72]. AIs have been developed for applications ranging from emotional learning companions that support elementary school children learning to code [124] to discussing Jane Austen books with lifelong learners [133]. AIs to support collaborative learning are underinvestigated, even though collaborative learning has been shown to increase learner interest in solving problems during online tutoring [7], decrease learner boredom [7], and improve critical thinking skills [59]. Researchers have begun to uncover features of strong collaboration, including gaze synchronization [155], the importance of proximity [25], and semantic similarity [149]. Substantial work has investigated collaborative processes with a

This chapter is based on a previous publication: Amanda E. Griffith, Gloria Ashiya Katuka, Joseph B. Wiggins, Kristy Elizabeth Boyer, Jason Freeman, Brian Magerko, Tom McKlin. “Discovering Co-creative Dialogue States during Collaborative Learning.” *Proceedings of the International Conference on Artificial Intelligence in Education (AIED)*, 2021, pp. 165-177.

- 1 This work is supported by the National Science Foundation through the grants DRL-1814083 and DRL-1813740. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.
- 2 EarSketch is available online at <https://ears sketch.gatech.edu>

variety of data sources including eye tracking [39][155], motion sensors [25][137][39], dialogue analysis [19][149], and speech features [187].

Research on the dialogue of collaboration has focused on detecting when students are off-task [19], supporting inquiry learning by analyzing the role of questions in collaborative computational modeling [167], [167] and predicting problem-solving modes to support adaptive tutoring [142]. Research on collaborative learning within groups has used conversational agents to facilitate productive conversations [43] and dialogue features to identify trouble during collaboration [60].

Most research on virtual agents in collaborative learning has involved agents in a tutor or support role, but some work with agents as partners to human learners has demonstrated benefits including significantly higher levels of shared understanding, progress monitoring, and feedback [146]. Research on agents as partners has also investigated support for human-computer co-creation, a type of collaborative creativity in which responsibility for an artifact is shared between the human and computer [83]. To move toward systems that support collaborative co-creation during learning, we need to build an understanding of the dialogue mechanisms that characterize this process.

To address this need, this work makes a step toward characterizing the dialogue modes learners tend to enter as they engage in co-creative dialogue. We examined dialogue and system interactions between pairs of high school students learning to code through remixing musical samples. Using a hidden Markov model, we distinguished seven states, three characterized by conversation and four characterized by task actions. The model suggests that learners engaged in two types of discourse—aesthetic and technical—during this co-creative process. The *aesthetic* discourse pertains to musical style, taste, and expression, while the *technical* discourse pertains to writing code and task objectives. By modeling co-creative dialogue, we can move toward intelligent support of human collaboration and toward intelligent co-creative agents that support learning.

This work examines the following research question: *RQ1: What are the dialogue states that characterize the co-creative process between learners, and how do learners transition between these dialogue states?*

3.2 System Description

The EarSketch web client contains a Code Editor, Digital Audio Workstation (DAW), Content Manager, and Curriculum (see Figure 3-1). Users create Python or JavaScript projects and edit them in the Code Editor, using EarSketch API functions such as `fitMedia()` and `makeBeat()` to arrange sound samples onto audio tracks for a given duration and apply audio effects. Once they compile the code, the DAW displays the result and the user can play, pause, loop, and solo or mute individual tracks to hear their project.

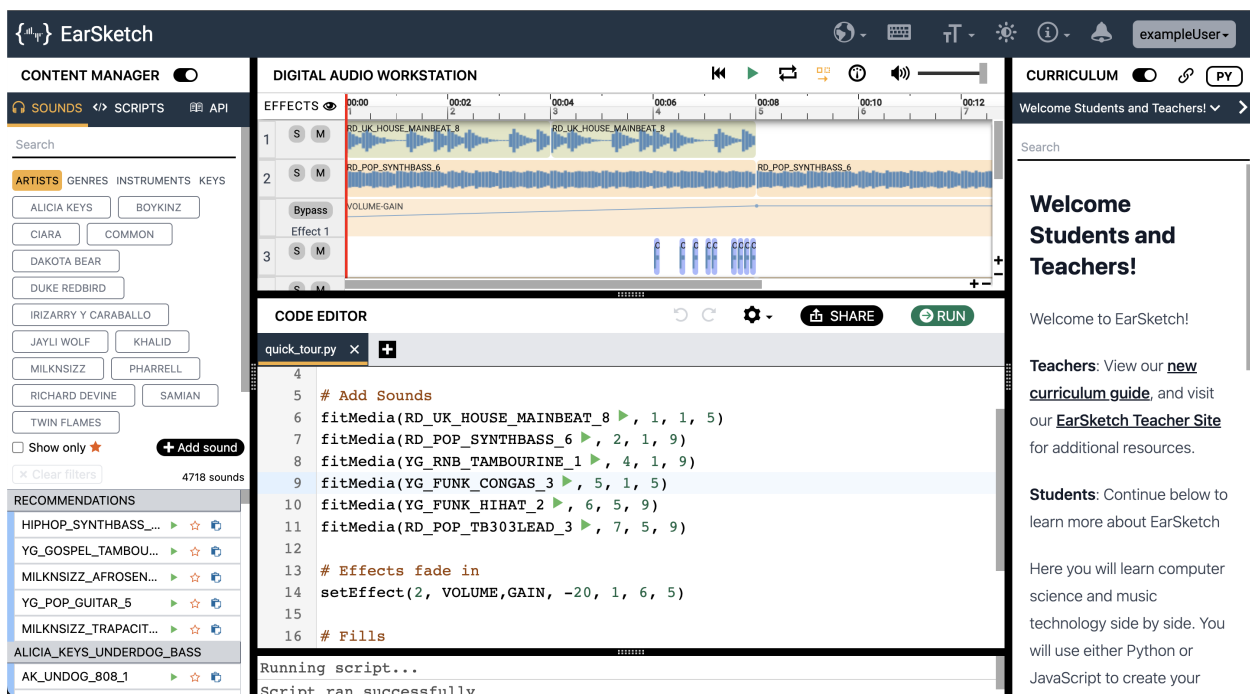


Figure 3-1. Primary components of the EarSketch web-based development environment: Content Manager with sound browser and project management tools (left), Digital Audio Workstation for audio visualization and control (top-center), Code Editor for Python and JavaScript implementation (bottom-center), and Curriculum pane containing instructional materials (right).

Users can search for sounds to use, manage the multiple projects they have created, and search for EarSketch API functions and documentation in the Content Manager. The Sound

Browser contains a text search as well as genre, artist, instrument, and key signature filters. Additionally, a recommendation system uses the contents of an active project to present complementary sounds to a user based on audio analysis as well as combinations of sounds used by others [165, 163, 166].

The Curriculum pane on the right side of the interface contains lessons and example code for various coding and music concepts such as loops, function definitions, musical form, debugging, and peer review. These lessons have been designed to be integrated with a full computer science curriculum, as well as the “Hour of Code” module for a single-session introduction to creative computing. Additionally, the EarSketch site has been the stage for multiple Your Voice Is Power competitions in partnership with Amazon Future Engineer³ designed to increase student involvement in CS education with themes of social change and equity.

3.3 Methods

This work analyzes a corpus of textual student-student dialogue collected between November 2019 and March 2020 during computer science classes from eight public high schools in two districts in the southern United States, consisting of a total of 140 participants. More than half of the schools had a student population of majority (>50%) Caucasian students; one school was majority Black; two schools had a substantial (>25%-35%) Latinx population; one school had a substantial Asian population. All students were in grades 10-12 (15-18 years old). Teachers placed the students into pairs, and students were placed at a distance in separate rooms or different areas of the same room to facilitate their communication through the textual chat interface (Figure 3-2). Students collaborated synchronously for an average of 48 minutes to remix musical samples and create an original song or ringtone. Some participants, 9 pairs, split their work across two class days. We included only the first day’s dialogue for these pairs because concatenating two separate dialogues would change the natural beginning, middle, and end of the sequences; whereas including both dialogues for a pair would unevenly weight their patterns while training the hidden Markov models.

3 <https://www.amazonfutureengineer.com/yourvoiceispower>

3.3.1 Learning Environment for Computational Music Remixing

This study was conducted in the EarSketch learning environment, an online interface for developing computational music (Figure 3-2). In prior studies, students that used EarSketch had significant positive results related to content knowledge and attitudes towards computing, especially in currently underrepresented groups in computing [111]. The EarSketch interface includes a code editor for Python or JavaScript and a digital audio workstation that allows users to access the music they have written [50]. The interface also features a content manager with samples (sound clips) that can be used to create music, as well as a curriculum tab that provides helpful resources associated with the class. Both students had access to all of the tools allowing both to contribute to the code simultaneously. In this study, the interface included a chat box to communicate with their partner. We logged all students' textual dialogue, all changes made in the code editor, all items accessed in the curriculum tab, and the results of the students running their scripts (such as successes or errors).

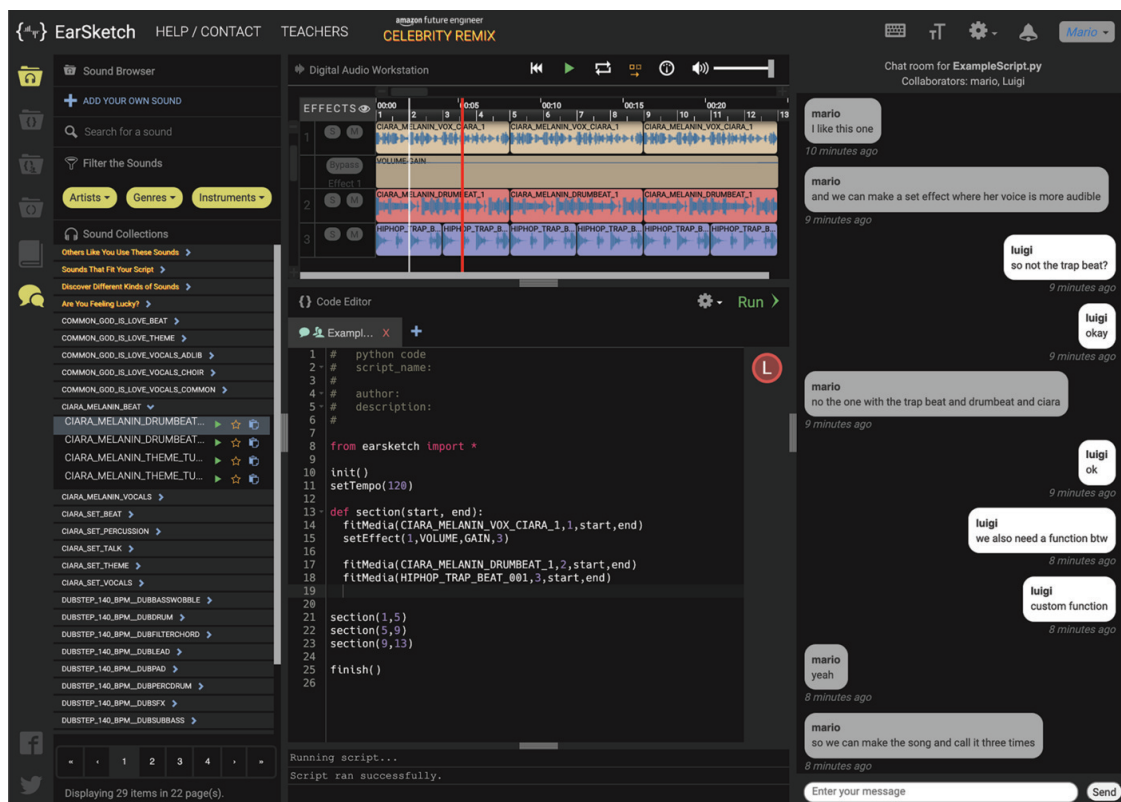


Figure 3-2. The modified EarSketch environment with chat window (right) used during the study.

3.3.2 Dialogue Tagging

In cleaning the dataset, we removed two sessions that contained exclusively off-task, joking, offensive, or gibberish content. The remaining textual dialogue corpus includes 68 sessions (136 students) and 3305 utterances, with a mean of 48 utterances per session ($SD=35$, $Min=4$, and $Max=214$). We developed and applied a dialogue act taxonomy that included 16 labels, which three independent annotators applied with a kappa of 0.76, *substantial*, agreement [103]. Among the 16 original labels, 10 occurred with greater than 5% probability in the hidden states within the HMM reported here, and one more label appeared in an example excerpt in this paper’s discussion section. These 11 relevant dialogue act labels are shown in Table A-1.

3.3.3 Hidden Markov Model

We implemented a hidden Markov model (HMM) to analyze the learners’ interactions and model the co-creative sequences [136] after compiling the lists of sequential observation symbols that represent the collaborative interactions. We chose to use an HMM because we were interested in the hidden discourse states of co-creative dialogue.

In an HMM, sequences of *observation symbols* represent observable events such as textual messages and coding actions. In Figure 3-3 the circles with the dialogue act labels represent the observation symbols. Hidden states are influences upon those observation sequences. Each hidden state is characterized by an *emission distribution*, which is a probability distribution over observation symbols. When the model is learned, every observation is modeled as having been “generated” by a hidden state. In Figure 3-3, the hidden states are represented by the unlabeled squares, and the arrow from the hidden states to the observations indicates the relationship of the observation having been generated by the hidden state.

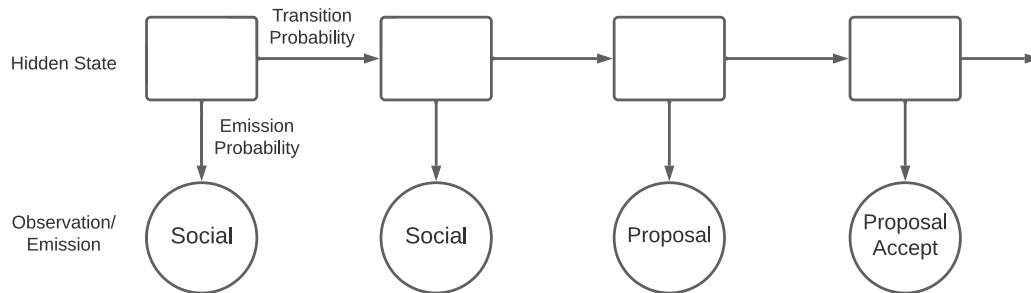


Figure 3-3. Example HMM Diagram.

Each hidden state also has a set of transition probabilities and an initial state probability. The transition probabilities indicate how likely the model is to either continue in that state or transition to another state are represented by the arrows from one hidden state to the next in Figure 3-3. The initial state probability that indicates the likelihood of the Markov chain beginning in that state. The HMM also has a stationary distribution that indicates the proportion of interactions that occur in each state over the long run. By training an HMM on the data from this study, we can take a high-level view of the learners' interactions and uncover generalizable patterns in the co-creative process.

The labeled dialogue and task actions are the observation symbols in this model. There are 20 distinct possible observation symbols—16 dialogue acts and the following actions:

- **curriculum:** The student accesses the curriculum or moves between lessons.
- **edit:** Any consecutive insertion or removal of characters in the code editor.
- **success:** Each time the script runs successfully.
- **error:** Each time an error is received when the script is run.

We represented each of the 68 collaborative dialogues as a sequence of these observation symbols and trained an HMM on these sequences. We did not model time between actions, nor did we model which of the two students performed each action.

3.4 Results

We used a hidden Markov model to answer *RQ: What are the dialogue states that characterize the co-creative process between learners, and how do learners transition between*

these dialogue states? One of the major decisions when learning an HMM is choosing how many hidden states to include. Sometimes the number of hidden states is known or strongly implied by physical or theoretical constraints of an application, but in our case, we had no prior notion of how many co-creative dialogue states would emerge. To identify the best number of hidden states n , we used leave-one-out cross-validation and compared the average Akaike information criterion (AIC) score for each number of hidden states [1]. We compared models using $n = 4$ to $n = 9$ states, finding the best AIC scores with $n = 6$ and $n = 7$. We then trained ten models each for both six and seven states and picked the single best model for each n using log-likelihood. The best models for each n were nearly identical: one of the dialogue states for the six-state model splits into two dialogue states in the seven-state model, with the other states remaining the same. We opted to move forward with the 7-state HMM.

The HMM analysis suggests seven hidden states within the co-creative dialogues (see Figure B-1). These hidden states are characterized by emission probability distributions over the observation symbols and must be interpreted as a whole. We interpret the learned hidden states as follows:

- **Social Dialogue:** In this state, *social* dialogue acts are emitted with a high probability (79%). Due mainly to greetings and other social dialogue at the outset of collaboration, around 90% of sessions begin in this state.
- **Aesthetic Dialogue:** This state is characterized by the observation symbols *proposal* (34%) and *proposal acceptance* (17%), which are the dialogue acts involving assertions and acceptances of creativity. The dialogue that belongs to this state usually involves discussing some aspect of the music.
- **Curriculum Browsing:** In this state, the observation symbol *curriculum*, which involves accessing the curriculum documents, is emitted with a very high probability (91%). A student browsing through several curriculum documents rather than accessing specific relevant documents characterizes this state.
- **Code Editing:** The observation symbols from this state are almost entirely code editing, which is emitted with a 99% probability.
- **Technical Dialogue** The observation symbols that characterized this state involved dialogue acts of *statement* (25%), *directive* (15%), *confusion* (13%), *acknowledgment* (9%), and *directive acceptance* (6%). The dialogue that belongs to this state usually involves discussion of code features or task requirements.

- **Compilation Success:** Mostly characterized by students running the code successfully, this state involves some *positive feedback* (6%) and *statement* (5%) dialogue acts.
- **Compilation Error:** Mostly characterized by students receiving an error when running the code, this state also involves some *confusion* (8%), *statement*(5%), and *directive*(5%) dialogue acts.

This model revealed three distinct states of conversation in these co-creative interactions: *Social Dialogue*, *Aesthetic Dialogue*, and *Technical Dialogue*. The *Social Dialogue* state usually occurs at the beginning of the interaction, but can occur throughout and usually includes some rapport building and off-task discussions. Utterances in the *Aesthetic Dialogue* state usually involve discussing different aspects of the music such as instruments, tempo, genre, and even what artist to emulate. Utterances in the *Technical Dialogue* state are typically about task requirements and code. This model also reveals four hidden states focused on coding: *Curriculum Browsing*, *Code Editing*, *Compilation Success*, and *Compilation Error*. The sessions never begin in the *Curriculum Browsing* state, and no other states consistently lead to it. Every state, excluding the *Social Dialogue* state, has a significant chance to lead to the *Code Editing* state, and 21.3% of the actions occur in the *Code Editing* state. *Code Editing* transitions to a *Compilation Success* or a *Compilation Error* state with a probability of 98.5%. The *Compilation Success* state is likely to transition to *Code Editing* (42.7%), *Technical Dialogue* (9.89%), and *Aesthetic Dialogue* (6.17%). The *Compilation Error* state is likely to transition to is *Code Editing* (68.8%), and then *Technical Dialogue* (3.60%).

3.5 Discussion

3.5.1 Dialogue States

The HMM reveals how co-creative pairs move among collaborative dialogue states characterized by conversation and task actions. Of the seven hidden states identified by the HMM, three are composed primarily of dialogue acts. *Social Dialogue* is the most likely state for pairs to start in, primarily composed of greetings and off-task dialogue. This is a typical feature of collaborative dialogue, prefacing discussion with periods of rapport building in which the partners become more familiar with each other [131]. After leaving this initial *Social Dialogue* state, the

conversation is nearly twice as likely to move directly to the *Aesthetic Dialogue* state (60%) as the *Technical Dialogue* state (31%). In the *Aesthetic Dialogue* state, pairs brainstorm and exchange dialogue related to the musical aspects of their co-creative task, such as genre and types of sounds. The *Technical Dialogue* is where pairs begin planning how to accomplish their aesthetic goals.

In the excerpt in Table 3-1, the pair sets their goal of creating a “dubstep” song and then debates how they would accomplish that in their code. This transition from *Aesthetic Dialogue* → *Technical Dialogue* accounts for 13.0% of the transitions across the sessions. In contrast, the transition from *Technical Dialogue* → *Aesthetic Dialogue* is less than half that at 6.02%. This observation suggests that most pairs tend to decide on what they want to make before they move on to making it.

Table 3-1. Student collaborative dialogue excerpt demonstrating the transition from *Aesthetic Dialogue* → *Technical Dialogue*. Each dialogue state was determined automatically using the HMM presented in this work.

State	Action	User	Text
Aesthetic	Passing Responsibility	Student 1	what you want to do
Aesthetic	Proposal	Student 2	lets do dubstep cause its fire
Aesthetic	Proposal Accept	Student 1	i feel you
Technical	Confusion	Student 1	what did she say how many variables
Technical	Statement	Student 2	3
Technical	Directive	Student 1	ok lets do this
Technical	Confusion	Student 2	ngl im kinda lost already so im sorry
Technical	Confusion	Student 1	i dont know what to do
Technical	Confusion	Student 2	me either ima ask for help

We use a ridgeline plot (Figure 3-4) to visualize where the hidden states occur in the sessions. The highest points are where the interactions in each state occur most frequently but are not relative to the other states’ frequencies. The pattern we observe using all 68 sessions is that the highest frequency of *Social Dialogue* occurred at the beginning, followed by the highest frequency of *Aesthetic Dialogue*. The highest frequency of *Technical Dialogue* occurs just before the editing and debugging process begins, but *Technical Dialogue* is more consistent throughout the sessions than *Social Dialogue* or *Aesthetic Dialogue*. Overall, dialogue appears to be more frequent at the beginning of the sessions and decreases as the editing and debugging processes take precedence.

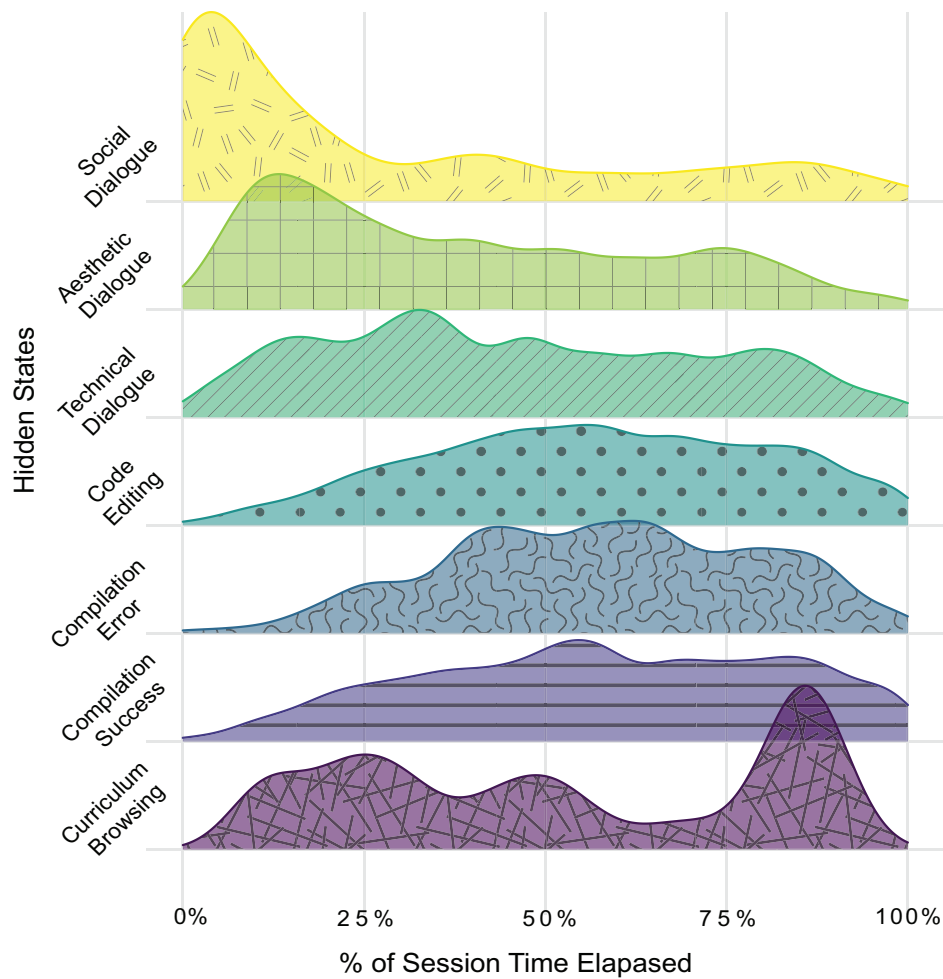


Figure 3-4. Hidden states as they occur throughout all 68 sessions. The widest part of each curve corresponds to the most frequent occurrence of that state during the sessions. This frequency is not relative to other states' frequencies.

3.5.2 The Debugging Process and the Conversation it Inspires

The remaining four states are primarily composed of actions in the interface:

- Browsing curriculum documents (*Curriculum Browsing* state)
- Working in the code editor (*Code Editing* state)
- Encountering compilation errors (*Compilation Error* state)
- Successfully compiling code (*Compilation Success* state)

The transitions between the *Code Editing*, *Compilation Success*, and *Compilation Error* states demonstrate the movement between collaborative states that occur during debugging, and

they offer insights into how co-creative conversations unfold around debugging. The *Code Editing* state primarily transitions to the *Compilation* states (*Error* or *Success*), with only four transitions to the *Social Dialogue* state in the entire dataset and zero transitions to the *Aesthetic Dialogue* or *Technical Dialogue* states. After the *Compilation Success* state, students are most likely to transition back into the *Code Editing* (43%) state, but they sometimes transition back into the *Technical Dialogue* (10%) or *Aesthetic Dialogue* (6%) states. The excerpt in Table 3-2, depicts a pair successfully compiling their code and transitioning to the *Technical Dialogue* state, discussing the measure lengths. The pair then transitions to the *Aesthetic Dialogue* state, making proposals about what to do with sounds and overall song length. Later in the session, after another successful code compilation, the pair transitions straight to *Aesthetic Dialogue*, discussing other sounds to add to their artifact.

Table 3-2. Student collaborative dialogue excerpt of dialogue sequences following successful code compilation. Each dialogue state was determined automatically using the HMM presented in this work.

State	Action	User	Text
Success	Compile	Student 1	Success
Technical	Acknowledgment	Student 2	oh
Technical	Statement	Student 1	ohh its because they are 2 second each
Technical	Statement	Student 2	each measure isnt exactly 2 seconds
Technical	Statement	Student 2	its a little longer i think
Technical	Statement	Student 1	but when another is added to the 30 it becomes exactly 2 seconds longer
Technical	Statement	Student 2	measure 15 is at 28.5 seconds
Technical	Compile	Student 2	Success
Aesthetic	Proposal	Student 1	we can use a combination of sounds
Aesthetic	Proposal	Student 2	we should just leave it at 31
Aesthetic	Proposal	Student 2	i think that will be fine
Aesthetic	Proposal Accept	Student 1	yeah
...	...	...	...
Success	Compile	Student 2	Success
Aesthetic	Editing Code	Student 1	...
Aesthetic	Proposal	Student 2	we should put like a synth or something to that effect
Aesthetic	Proposal	Student 2	add some like futuristic noises or airhorns or something

The *Compilation Success* state seems to be an inflection point in the co-creative process, in which the pairs may start to renegotiate some of the creative aspects of their code. In contrast, the *Compilation Error* state only transitions to *Code Editing* or *Technical Dialogue* states, suggesting partners who encounter errors focus on resolving their problems rather than discussing the aesthetic elements of the artifact.

3.6 Key Takeaways for My Dissertation Research

This chapter overviews my work which investigates the textual dialogue and system interactions of high school pairs to learn a hidden Markov model so we can identify the co-creative states and common transitions between the states. Three states are associated with dialogue—Aesthetic, Technical, and Social—and four are associated with interface actions—Curriculum, Code Edit, Compilation Success, and Compilation Error. Common patterns include the beginning of the sessions being very dialogue-heavy, with pairs typically beginning with greetings and rapport building (Social Dialogue state), followed by discussing what the pairs would like to make (Aesthetic Dialogue state). After the pairs decide what the pairs want to make, the pairs transition to discussing the task and technical requirements (Technical Dialogue state) before finally moving on to implementing their ideas. Section 3.5.2 reveals the dialogue patterns observed in the debugging process. These dialogue patterns suggest the Compilation Success state served as an inflection point in the co-creative process, where the pairs renegotiate some of the aesthetic aspects of their artifact after the pairs begin coding.

These insights provide a foundational understanding of the interplay between dialogue and action within a co-creative partnership. The states associated with dialogue highlight the importance of communication in shaping the co-creative process, suggesting potential points where conversational AI could offer meaningful support. However, the mere identification of common patterns does not implicitly confirm their efficacy or relationship with successful outcomes. It still remains to be seen, based solely on this study, how these patterns correlate with measures of creativity, learning, or satisfaction with the co-created product. Therefore, in my

followup analysis (Chapter 4) I examine the relationship between the co-creative states and the outcome of partner satisfaction.

CHAPTER 4

INVESTIGATING THE RELATIONSHIP BETWEEN DIALOGUE STATES AND PARTNER SATISFACTION DURING CO-CREATIVE LEARNING TASKS

This chapter presents a study that builds on the work from Chapter 3, where I transition from identifying the co-creative states and common state transition patterns to examining their association with the students' satisfaction with their collaborative partners. The insights gained shed light on the relationships of certain dialogue dynamics—especially those involving the navigation of technical and creative challenges—with the partners' satisfaction with their collaborative efforts.

4.1 Background and Motivation

Collaborative learning, in which two learners work together on a shared problem or goal, offers many benefits: it has been shown to increase learner interest in solving problems during online tutoring [7], decrease attrition rate in MOOCs [47], and improve critical thinking skills [59]. The interactions that occur during collaborative learning can be observed in many ways, such as gaze synchronization [155, 39]; body proximity and posture [137, 39]; and dialogue [149, 19, 167], which plays a fundamental role. Dialogue mediates joint problem solving and knowledge building by allowing collaborators to draw attention to problems and propose solutions [173]. Research into collaborative dialogue has focused on detecting when students are off-task [19], analyzing the role of questions in collaborative computational modeling [167], predicting problem-solving modes [142], and identifying issues during collaboration using dialogue features [60].

While many forms of collaborative dialogue have been well studied, co-creative dialogues—collaborative dialogues that occur during co-creative activities—are understudied. Within co-creative activities, people, or people and machines, work together and share responsibility for creating an artifact [58, 83]. How key dialogue phenomena manifest within

This chapter is based on a previous publication: Amanda E. Griffith, Gloria Ashiya Katuka, Joseph B. Wiggins, Kristy Elizabeth Boyer, Jason Freeman, Brian Magerko, Tom McKlin. “Investigating the Relationship Between Dialogue States and Partner Satisfaction During Co-Creative Learning Tasks.” *International Journal of Artificial Intelligence in Education*, 2022.

co-creative dialogues for learning is an open research question, and understanding these processes is an essential step toward supporting collaborative co-creation in education.

To address this need, this chapter investigates co-creative dialogues in the domain of computational music remixing. Specifically, we examine dialogue and system interactions between pairs of high school students learning to code through remixing musical samples. Studying the collaborative dialogues of pairs as the pairs engage in computational music remixing provides insight into the unique exchanges that occur as the pairs negotiate aesthetic decisions while navigating technical constraints. This work examines the following research question in the context of computational music remixing with high school learners: *What are the relationships between the outcome of partner satisfaction and the dialogue that comprise the co-creative process?*

To investigate this question, we first labeled the dialogue with dialogue acts (Table A-1) and then modeled dialogue states with a hidden Markov model (HMM) over those sequences of dialogue acts interleaved with events in the computational music remixing interface, as described in Chapter 3. After learning the seven-state HMM, we used the relative frequency of each pair's states and their transitions to conduct a regression analysis to predict the partner satisfaction ratings from the post-survey.

The results from this analysis reveal that the relative frequency of actions in certain states and some transitions between states were predictive of partner satisfaction. We found that partner satisfaction is negatively associated with the relative frequency of the *Compilation Error* state and with the relative frequency of transitions from *Curriculum Browsing* → *Code Editing*. Partner satisfaction is also negatively associated with the relative frequency of transitions from *Aesthetic Dialogue* → *Technical Dialogue* and *Aesthetic Dialogue* → *Code Editing*.

This work is the first to model the dialogue mechanisms of high school learners' co-creative interactions and provide insights into the relationship between these interactions and partner satisfaction. By modeling co-creative interactions and their relationship with partner satisfaction, we can move toward intelligent support of co-creative learning. In the following sections, we

present related work, describe our methodological approach, discuss the relationship of the hidden states and their transitions with partner satisfaction, and discuss implications for designing intelligent support systems.

4.2 Related Work

Collaborative learning is a complex process involving both cognitive and social dimensions. Collaborative learning is broadly defined as “a situation in which two or more people learn or attempt to learn something together” [40]. While there are a variety of more explicit definitions for collaborative learning, most tend to include some form of *creation of shared knowledge by groups while working together towards a common goal* [34, 145]. The notion of working together towards a common goal is also part of collaborative problem solving. Research provides nuanced descriptions of the features that distinguish collaborative problem solving from other forms of collaboration which include solving a novel problem, objective accountability of solution quality during problem solving, differentiation of roles, and interdependency among team members [62].

While focus on collaborative learning has been increasing, not enough is understood about it in K-12 contexts. Specifically, while practices such as collaborative computing are identified as core practices in the US’s K-12 CS Framework [100], most research into collaborative programming happens in undergraduate computer science courses. Research shows that undergraduate learners who participate in collaborative programming produce higher quality code, greater confidence in their work, and are more likely to pass their programming courses [16, 116]. Research into K-12 collaborative programming practices includes comparing the benefits of pair programming for middle school students to working individually [38], exploring the different ways pairs’ dialogue, suggestions, and roles affect collaboration [182, 183], asking how attitude toward collaboration and prior knowledge of programming influences interactions [18], and investigating how pair programming improves confidence and performance in primary school girls [203].

Another practice identified in the US’s K-12 CS Framework is creating computational artifacts, which is a process that “embraces both creative expression and the exploration of ideas to create prototypes and solve computational problems. Students create artifacts that are

personally relevant or beneficial to their community and beyond.” [100] Studies consistently find benefits to incorporating more creative activities into computer science classes, such as increased motivation, engagement, and content knowledge.

Research identifies two types of students interested in computer science and programming [95]. The first type views computer science as “fun, creative, and autonomous”, and the second type is more interested in acquiring knowledge and solving problems. The first group finds their computer science classes disappointing, suggesting that more opportunities for creativity in computer science courses may be necessary to prevent loss of interest. A study on the effects of computational music remixing on attitudes and knowledge of computer science on high school students finds students value these creative and personally relevant aspects, and results suggest that these aspects may be even more engaging for female and minority students [49]. These findings are supported by additional research which observes that while all groups of students show significant increases in intent to persist in computing and content knowledge, the increase in female student interest in computing is significantly higher than the increase in male interest [111].

Despite the benefits of collaboration and creativity in computer science education, there is little research into the processes of co-creativity in computer science classrooms. A study exploring co-creative processes and how they impact concept formation finds that, compared to non-interactive and interactive non-co-creative processes, the co-creative process enhances concept formation depth but also requires more time and a greater cognitive load [115]. This finding indicates that we cannot assume the same processes are involved between collaborative programming and co-creative programming. In fact, research on phases of collaboration among different activities notes how the phase order in “certain types of creative problems” is the reverse of what was predicted, and that these tasks may require a “different sort of phase hypothesis” [10]. When designing systems to support co-creative learning, it is essential to understand the co-creative processes that occur and how they might impact learner outcomes.

There is a growing interest in investigating partner satisfaction rather than focusing exclusively on individual outcomes in co-creative domains. Historically, partner satisfaction has

been seen as important in reflecting the extent to which collaborators are engaged or adding value to the collaboration [188]. Recent work supports this finding: students' attitudes about each other are found to influence the extent to which they contribute during pair programming [18]. Similarly, dialogue acts reflecting higher participation are associated with higher partner satisfaction during pair programming activities [87]. Furthermore, there seems to be a relationship between partner satisfaction and how student partners talk with one another [88]. However, little is known about the relationship between partner satisfaction and higher-level processes in co-creative group interactions.

Hidden Markov models (HMM) are well-suited to learning higher-level constructs from observable interactions during collaboration in computer science. Research comparing five computational approaches—finite state machines, rule learners, decision trees, plan recognition, and hidden Markov models—for modeling knowledge sharing during collaborative learning finds that the HMM performs significantly better than the other approaches [169]. In the domain of computer science learning, an HMM learned from a corpus of human-human introductory computer science tutoring dialogues [14] demonstrates that an HMM can be used to identify tutorial dialogue strategies from the structure of the dialogue act sequences. Further research using both dialogue acts and task actions from human-human tutoring demonstrates that HMMs can identify meaningful high-level structures (i.e., hidden states) that are correlated with learning outcomes when the low level codes (i.e., observation symbols) that make up these states are not [15]. In addition to computer science tutoring, this method has also been applied to collaborative programming processes. A study using HMMs to compare the problem-solving approaches of individual students and pairs of students during programming activities highlights the similarities and differences of the types of states each moves through and where the pairs employ different problem-solving patterns and strategies than the individuals [142]. For instance, only pairs have the *program planning* and *program building* states. Researchers note that the *program planning* state might be most representative of the collaborative aspect of the task. Additionally, they note that pairs are more likely to persist in the same problem-solving state for longer than individuals.

In summary, the prior research has established that in CS education, collaboration is beneficial, creative coding can help motivate and engage students, and HMMs can be used to identify the higher-level processes of group interactions. The work presented in this chapter builds on previous research to identify the group processes of collaboration and creativity in learning. While earlier research has focused on exploring these processes separately or in other learning domains, we focus on learning the processes of co-creative learning in K-12 computer science education using an HMM. Additionally, while most research into collaborative learning focuses on learning outcomes or performance on an activity or artifact, we explore the relationship between co-creative processes and partner satisfaction.

4.3 Methods

This work analyzes a corpus collected from a study conducted in high school computer science classes. The corpus consists of textual student-student dialogue gathered while students worked together in pairs within the EarSketch.¹ learning environment, an online interface for developing computational music (Figure 3-2) In prior studies, students who used EarSketch, especially those from groups currently underrepresented in computing, have had significantly positive results related to content knowledge and attitudes towards computing [111]. The EarSketch interface includes a code editor and a digital audio workstation that allows users to access the music they have written [50]. The EarSketch interface also features a content manager with samples, which students can use to create music. Additionally, a curriculum tab provides helpful resources associated with the class.

During the study, both students in each pair had access to all tools, allowing both to contribute to the code simultaneously. Additionally, the interface included a chat box for partners to communicate with one another. We logged all pairs' textual dialogue, the activity in the code editor, the items viewed in the curriculum tab, and the outcome of the students running their scripts (e.g., successful compile or error received).

1 EarSketch is available online at <https://earsketch.gatech.edu>

This study analyzes dialogue data collected from 140 participants across eight public high schools in two districts in the southern United States. More than half of the schools have a student population of majority (>50%) White students; one school is majority Black; two schools have a substantial (>25%-35%) Latinx population; and one school has a substantial Asian population. Students were in grades 10-12, ranging from 15-18 years old.

To encourage the students to communicate through the textual chat interface (Figure 3-2), teachers placed the pairs either in separate rooms or different areas of the same room. Students collaborated synchronously for an average of 48 minutes to remix musical samples and create an original song or ringtone. Nine pairs split their work across two class days. We only included the first day's dialogue because including the dialogue from both days would unevenly weight the patterns of those nine pairs while training the hidden Markov models while concatenating the dialogues from the two separate days would change the natural beginning, middle, and end of the sequences.

After their collaboration, students were given a post-survey about their satisfaction with their partner, which we will discuss further in Section 3.3. Some students were also given a pre-survey asking students to describe their experience, confidence, and enjoyment in both computing and music. However, because the pre-surveys were not administered during the first sets of classroom studies, we only have pre-survey data for 44 of the 136 students. Therefore, we did not include the pre-survey data in our analyses.

4.3.1 Partner Satisfaction Survey

Each participant filled out a post-survey where they rated their partner on seven items using a 4 point scale of Strongly Disagree (1) to Strongly Agree (4) with no neutral answer. Figure 4-1 shows the summary of these results. We conducted a Principal Component Analysis (PCA) on the survey item responses that we collected. The PCA results suggest that these seven items should be considered one PCA cluster, so we combined these seven items to obtain a single *partner satisfaction* for each student. We then summed the two *partner satisfaction* scores of the students in each pair to form a *combined satisfaction* score. The *combined satisfaction* score provided a

group metric that indicates overall partner satisfaction. The range for possible *combined satisfaction* scores is from 14 to 56, with actual scores that ranged from 19 to 56 ($M = 46.4$, $SD = 6.17$).

The post-survey also asked students to assess their partner’s coding knowledge in response to the prompt *My partner’s coding knowledge was:* with a scale of *Much less than mine*, *Somewhat less than mine*, *About the same as mine*, *Somewhat better than mine*, and *Much better than mine*. The post-survey also included two open response questions. The first was a follow-up to the *My partner made valuable contributions* item, —*If so, what was the most valuable contribution your partner made?* The second open response question was a follow-up to the *I would like to work with this partner again* item, —*How come?*

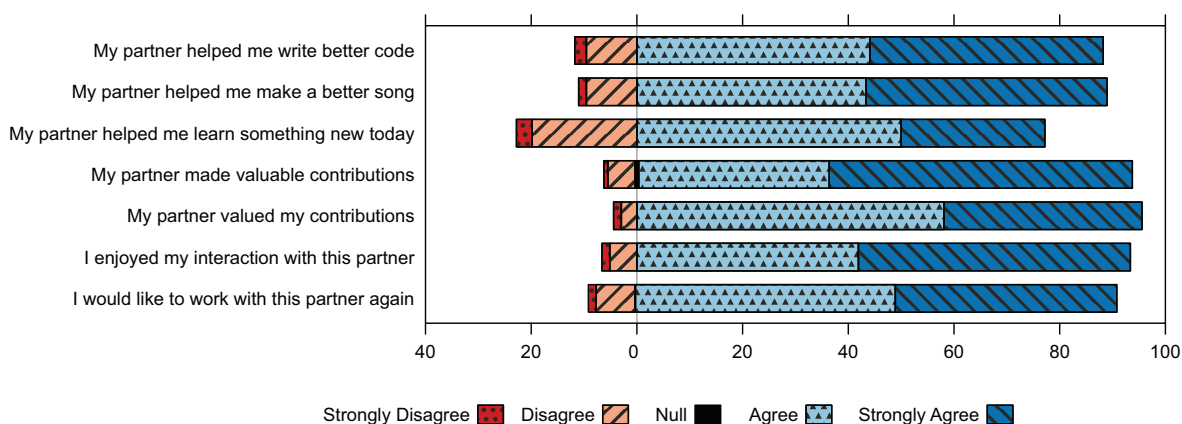


Figure 4-1. Post-Survey Responses on Perceived Partner Contribution

4.3.2 Using Hidden States and Transitions to Predict Combined Satisfaction

After learning an HMM (described in Chapter 3), our next step was to identify how the hidden states of the HMM (their relative frequencies out of the collaborative sessions, and the transitions between the hidden states during those collaborative sessions) were associated with the outcome of *combined satisfaction*.

To answer the question *What are the relationships between the outcome of partner satisfaction and the dialogue that comprised the co-creative process?* we built a predictive model

using these state and transition features as predictors. To identify which of the hidden states and their transitions predicted *combined satisfaction*, first we calculated the relative frequencies for each state and for the transitions among states.

To illustrate how we computed relative frequency, we use the example in Figure 4-2, which depicts an HMM that includes six observations, *Social*, *Social*, *Passing Responsibility*, *Proposal*, *Proposal Accept*, and *Statement*, and their hidden states, *Social Dialogue*, *Social Dialogue*, *Aesthetic Dialogue*, *Aesthetic Dialogue*, *Aesthetic Dialogue*, and *Technical Dialogue*. If these six collaborative moves represent a pair's entire session, the session has six observations, six (not necessarily distinct) hidden states, and five transitions. To compute the relative frequency for the *Social Dialogue* state, we count the number of times it occurs and divide by the total number of observations. In our simple example the model fits the *Social Dialogue* onto two of the six observation symbols, so the relative frequency of that state is $\frac{2}{6}$ or $\frac{1}{3}$. We compute this relative frequency for all states, so for our simple example $\frac{1}{2}$ of the interactions occur in the *Aesthetic Dialogue* state, and $\frac{1}{6}$ of the interactions occur in the *Technical Dialogue* state. Every other state in this example has a 0% for the relative frequency of interactions that occur in that state. This relative frequency is not equivalent to the elapsed time that is spent in each state as the HMM only considers the sequence order of the discrete observations and not the time duration.

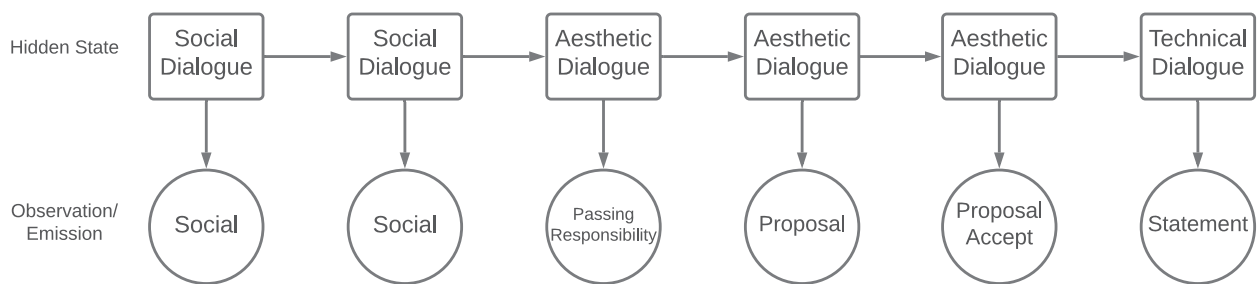


Figure 4-2. HMM transition example.

We used the relative frequencies of the transitions to investigate how the flow of the interaction was associated with *combined satisfaction*. To calculate the relative frequency of each transition, we divided the number of transitions from each state to the next state (e.g., the

transition from *Social Dialogue* → *Social Dialogue* or the transition from *Aesthetic Dialogue* → *Technical Dialogue*) by the total number of transitions in that collaborative session.

In our simple example, the transition from *Social Dialogue* → *Aesthetic Dialogue* occurs once out of the five transitions, so the relative frequency is $\frac{1}{5}$. Continuing this simple example, $\frac{1}{5}$ of the pair's transitions were from *Social Dialogue* → *Social Dialogue*, $\frac{2}{5}$ were from *Aesthetic Dialogue* → *Aesthetic Dialogue*, and $\frac{1}{5}$ were from *Aesthetic Dialogue* → *Technical Dialogue*. Any other possible transitions would have a 0% transition relative frequency.

The model assumes that the predictors are not highly correlated, but we discovered that there was a strong correlation between the relative frequency of interactions in a specific hidden state and the relative frequency of self-transitions of that hidden state. Therefore, prior to model building we removed self transitions for each state. For example, the relative frequency of interactions in the *Social Dialogue* state and the relative frequency of self-transitions from the *Social Dialogue* → *Social Dialogue* state had strong positive linear relationship ($r(68) = .97, p < .0001$), thus we removed the self-transition from *Social Dialogue* → *Social Dialogue* from the list of predictors. There were no self-transitions from the *Code Editing* state, so we chose to use the relative frequency of the hidden states instead of the hidden state self-transitions.

After computing the relative frequencies of states and state transitions for each pair's session, we treated those values as features within a regression model. We had 20 independent variables (predictors), which included the relative frequencies of the 7 hidden states and the relative frequencies of 13 transitions between the hidden states for each pair. The dependent variable was the *combined satisfaction* score for the pair. We used JMP [152], SAS's statistical analysis software suite, to conduct the regression analysis using best subset selection method to explore which of the relative frequencies of the hidden states and their transitions predicted the *combined satisfaction*.

The best subset selection method identifies the best subset of significant predictors. It considers all possible models from the null model, which includes only the intercept and no predictive features, to k predictors— k being the number of independent variables—and chooses

the best model for each size and then the best overall model from those models [151]. We interpret the best model, which we selected based on the Bayesian Information Criterion (BIC) [127].

The significant predictors of *combined satisfaction*, along with their estimated parameters, are shown in Table 4-1. For ease of interpretation, the predictors have been centered and scaled to provide standardized estimates so that we may compare the coefficients when we interpret the models.

The results show that the transitions from *Curriculum* → *Code Editing*, *Aesthetic Dialogue* → *Technical Dialogue*, and *Aesthetic Dialogue* → *Code Editing* have a negative relationship with *combined satisfaction*. This means that the larger the relative frequency of these transitions that pairs have, the lower the *combined satisfaction*. Additionally, the greater the relative frequency of the interactions that occur in the *Compilation Error* state, the lower the *combined satisfaction*.

Table 4-1. Statistical Analysis of Dialogue States and Transitions: Generalized regression results examining how dialogue states and state transitions relate to combined satisfaction scores ($n = 68$).

Variable	Coefficient	Std Error	p -value
Curriculum to Code Editing transition	-14.2	4.74	0.0026
Aesthetic Dialogue to Technical Dialogue transition	-20.5	8.81	0.0198
Aesthetic Dialogue to Code Editing transition	-15.2	6.88	0.0264
Relative frequency of interactions in Compilation Error	-21.4	8.93	0.0164

4.4 HMM's Relationship with Combined Satisfaction

4.4.1 Compilation Error State's Negative Relationship with Combined Satisfaction

The *Compilation Error* state has a negative relationship with *combined satisfaction*. The higher the relative frequency of interactions in the error state, the lower the *combined satisfaction* for the pair. Pairs that have more of their interactions in the *Compilation Error* state likely also have more issues completing their task, thus spending more time debugging. Prior research has established that, when compared to other programming tasks, students find the debugging process to be more tiring and less enjoyable when working in pairs [21].

We compared the top and bottom one-third of pairs based on their *combined satisfaction* to identify their distinguishing features. There were 21 pairs, around 31%, with *combined*

satisfaction greater-than-or-equal-to 50, and 20 pairs, around 29%, with *combined satisfaction* less-than-or-equal-to 44.

We use a ridgeline plot (Figure 4-3) to compare the pairs with the highest and lowest *combined satisfaction*. The pairs with the highest *combined satisfaction* have a single peak of *Compilation Error* state frequency around 60% through the session with most errors resolving by the end of their sessions. In contrast, the pairs with the lowest *combined satisfaction* have two peaks of *Compilation Error* state frequency and more unresolved errors at the end of their sessions.

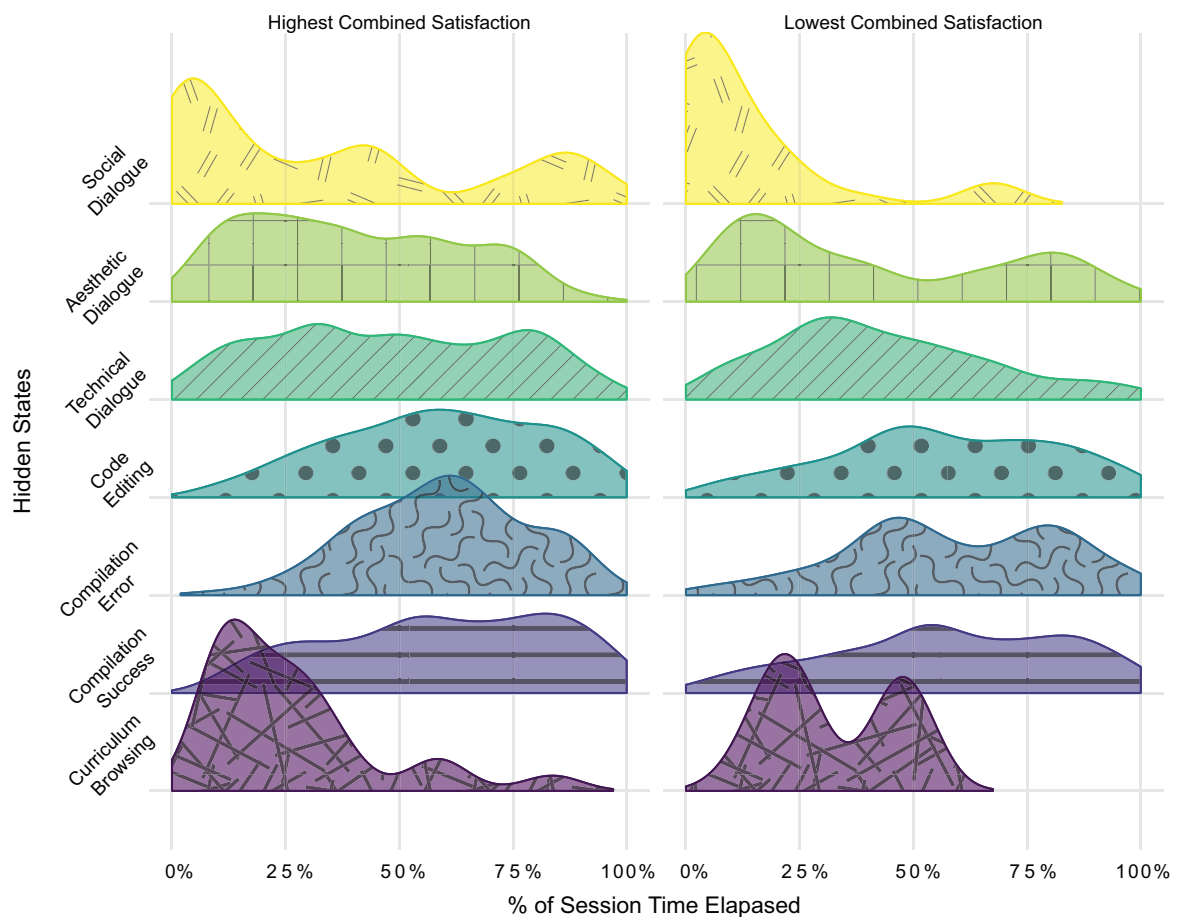


Figure 4-3. Hidden states as they occur throughout sessions for the 21 pairs with the highest *combined satisfaction* and 20 pairs with the lowest *combined satisfaction*. The widest part of each curve corresponds to the most frequent occurrence of that state during the sessions. This frequency is not relative to other states' frequencies.

An example of a lower rated pair that ends their session with compile errors and expressions of confusion can be seen in the excerpt in Table 4-2.

Table 4-2. Student collaborative dialogue excerpt of students ending with unresolved compilation errors and expressions of confusion. Each state was determined automatically using the HMM presented in this work.

State	Action	Student	Text
Error	Compile	Student 1	Unknown Identifier
Error	Passing Responsibility	Student 1	do you want to change anything to the song ?
Error	Passing Responsibility	Student 2	i'm not sure i have to listen to it so i can check
Error	Compile	Student 2	Unknown Identifier
Edit	Editing Code	Student 2	
Error	Compile	Student 2	Unknown Identifier
Error	Compile	Student 2	Unknown Identifier
Edit	Editing Code	Student 2	
Error	Compile	Student 1	Unknown Identifier
Error	Compile	Student 2	Unknown Identifier
Error	Compile	Student 1	Unknown Identifier
Technical	Confusion	Student 1	i am confused on what is going on
Technical	Confusion	Student 2	idk im so confused

Alternatively, in the excerpt in Table 4-3 we can see a pair with higher *combined satisfaction* ending the session with working code and positive feedback on their song.

Table 4-3. Student collaborative dialogue excerpt of students ending their session with no errors and demonstrating successful compilation and positive feedback. Each state was determined automatically using the HMM presented in this work.

State	Action	Student	Text
Success	Compile	Student 2	Success
Success	Statement	Student 1	i chose a random one i saw in the category
Edit	Code Editing	Student 2	
Success	Compile	Student 2	Success
Success	Compile	Student 1	Success
Success	Closing	Student 1	there we are done
Success	Positive Feedback	Student 2	the last guitar sounds pretty good

One of the main differences between the interactions of the pairs with the highest and lowest *combined satisfaction* is that the pairs with higher *combined satisfaction* tend to focus on solving errors together before moving on to other tasks and explicitly state when they are moving onto the next task once a task is completed. For example, the excerpt in Table 4-4, we can see students

trying to figure out why their song is not sounding the way they expected. *Student 1* suggests a solution, saying “we need to change the track,” and *Student 2* replies that “... so even if it was on the same track it should work.” They both edit the code, and *Student 2* comments on a change made by Student 1, saying “if we add it there we would need to add a parameter.” They continue to work on their artifact collaboratively, resolving one part of the issue and directing their attention to the next part with *Student 2* saying “Alright lets see where the problem is from here” and *Student 1* agreeing. The collaborative approach to solving errors from the previous excerpt can be contrasted with the earlier excerpt in Table 4-2, where *Student 2* tries to compile and listen to the song, but receives an error. Neither student mentions anything about the error until *Student 1* indirectly references it at the end of the session by saying “i am confused on what is going on” and *Student 2* agrees by saying “idk im so confused.” The lack of communication about code errors may be because when facing errors in their code, some students see the errors as a negative reflection of their abilities [92, 61]. The negative self-assessments related to errors in code could suggest that pairs that engaged in more discussion around errors were more confident in their abilities.

Additionally, expressions of uncertainty or confusion are an important aspect of collaborative dialogue and can provide learning opportunities when resolved [11]. However, unresolved uncertainty can lead to negative affective states such as boredom and frustration [44]. We find that higher rated pairs usually try to clear up any confusion about the task or artifact, even if the error has been resolved. For example, in the excerpt in Table 4-5, *Student 1* makes some changes to the code and resolving an error, but *Student 2* references the removed code saying “i was trying to make that an effect.” *Student 2* then makes some changes to the code and is able to successfully compile it. *Student 1* apologizes and indicates their confusion saying “my b i just didn’t know what was wrong.” Despite the problem being resolved and the code continuing to compile successfully after changes from Student 1, *Student 2* elaborates on their earlier explanation by replying “your all good, im trying to make the CIARA beat slower by setting an effect to change the tempo.”

Table 4-4. Student collaborative dialogue excerpt demonstrating focused problem-solving and explicit task progression. Each state was determined automatically using the HMM presented in this work.

State	Action	Student	Text
Technical	Statement	Student 1	we need to change the track
Technical	Editing Code	Student 1	
Technical	Statement	Student 2	But it goes from measures 1-8 and then 12-16 so even if it was on the same track it should work
Edit	Editing Code	Student 2	
Success	Compile	Student 2	Success
Edit	Editing Code	Both	
Error	Compile	Student 1	Type Error
Edit	Editing Code	Student 1	
Error	Statement	Student 2	if we add it there we would need to add a parameter
Edit	Editing Code	Student 2	
Error	Compile	Student 1	Type Error
Edit	Editing Code	Both	
Success	Compile	Student 2	Success
Edit	Editing Code	Both	
Error	Compile	Student 1	Type Error
Edit	Editing Code	Both	
Success	Compile	Student 2	Success
Technical	Directive	Student 2	Alright lets see where the problem is from here
Technical	Directive Accept	Student 1	ok

In contrast, the excerpt in Table 4-6 shows *Student 1* asking “do you know how to do the envelope?” and *Student 2* responds that they do not. *Student 1* returns to editing code, and the topic is never brought up again. If *Student 1* did find the solution, it is not discussed with their partner. This pair went on to only have six more utterances and ended the session with more unresolved expressions of confusion and compile errors. Research comparing uncertainty in pair programming for high and low performing pairs based on task solution quality find a similar pattern where higher performing pairs tend to address and resolve uncertainty and also focus on resolving a single task before moving on while lower performing pairs do not [143].

In our examination of the *Compilation Error* state, we find such communication between partners to be an important part of resolving errors, which is consistent with prior research. One study finds more types of transactive discussion—in which partners act on each other’s reasoning—in pairs and that the pairs generate a deeper understanding of the problem more

Table 4-5. Student collaborative dialogue excerpt of a student addressing their partner’s confusion and explicitly stating their goal. Each state was determined automatically using the HMM presented in this work.

State	Action	Student	Text
Error	Compile	Student 1	Type Error
Edit	Editing Code	Student 1	
Success	Compile	Student 1	Success
Success	Statement	Student 2	i was trying to make that an effect
Edit	Editing Code	Student 2	
Success	Compile	Student 2	Success
Success	Confusion	Student 1	my b i just didn’t know what was wrong
Success	Compile	Student 1	Success
Edit	Editing Code	Student 1	
Success	Compile	Student 1	Success
Success	Statement	Student 2	your all good, im trying to make the CIARA beat slower by setting an effect to change the tempo

Table 4-6. Student collaborative dialogue excerpt demonstrating unresolved confusion about technical implementation. Each state was determined automatically using the HMM presented in this work.

State	Action	Student	Text
Technical	Editing Code	Student 1	
Technical	Confusion	Student 1	do you know how to do the envelope?
Technical	Confusion	Student 2	i don’t
Technical	Editing Code	Student 1	

quickly [176] . Additionally, others observe improvement in the number of problems completed and debugging during pair programming with increased communication and certain types of transactive statements [125]. The increased communication in the pairs with the highest *combined satisfaction* may have helped the pairs resolve their errors more quickly, thus having fewer interactions in the *Compilation Error* state.

4.4.2 Curriculum Browsing → Code Editing: Negative Relationship with Combined Satisfaction

The model reveals that more transitions from *Curriculum Browsing* → *Code Editing* have a negative relationship with *combined satisfaction*. While examining the interactions of the pairs, we note two different ways—targeted and undirected—that students use the curriculum resource. The first type, targeted, is when students had a specific problem they were working on, like

resolving an error or writing a custom function, and they access the specific curriculum resources that addressed their problem. Students using the curriculum documents to address a specific problem is occasionally part of the *Curriculum Browsing* state, but it is most often included in one of the other states, such as the *Compilation Error* state. An excerpt of a student receiving a *Type Error* and then accessing the 31.7: *Type Error* curriculum document in the *Compilation Error* state can be seen in Table 4-7.

Table 4-7. Student collaborative dialogue excerpt presenting targeted curriculum resource usage as part of the Compilation Error state. Each state was determined automatically using the HMM presented in this work.

State	Action	Student	Text
Edit	Editing Code	Student 1	
Error	Compile	Student 1	Type Error
Error	Curriculum	Student 1	31.7: Type Error
Edit	Editing Code	Student 1	
Success	Compile	Student 1	Success

However, the type of undirected curriculum usage that characterizes the *Curriculum Browsing* state involves students clicking through the curriculum indiscriminately. Prior research has found that when students are not sure how to begin, they would freeze up or start looking through their resources for examples [92]. In some of the pairs with lower *combined satisfaction*, one student seems to be unsure of how to get started and does little or no coding within the session while their partner does either all or almost all of the coding. The students that browse the curriculum and make few contributions also ask questions like “how do I play it?” without any response from their partners. The curriculum browsing behavior usually starts from the introductory unit and follows it through the subsequent topics, which is depicted in the excerpt in Table 4-8. In the same session from the excerpt in Table 4-8, *Student 1* tries to add some code to the project about a quarter of the way through the session. *Student 2* says “What are you trying to write?” and deletes Student 1’s contribution. *Student 1* does not try to add anything else to the project for the rest of the session, but does go back to access some of the basic curriculum documents again. Another study has seen similar patterns where a higher skilled partner takes full control when paired with a more passive and lower skilled partner [21]. We can contrast this

behavior with an interaction from a pair with higher *combined satisfaction*. One student clicks through the curriculum at the very beginning of the session then asks their partner “*so we’re just making a ringtone.*” They continue to click through the curriculum for a couple of minutes until their partner answers “*yes*” and explains “*we need to figure out how many measures it takes to get to 60 seconds.*” After this explanation from their partner, the student stops browsing the curriculum and starts the project.

Table 4-8. Student collaborative dialogue excerpt demonstrating systematic curriculum browsing concurrent with partner’s code editing. Each state was determined automatically using the HMM presented in this work.

State	Action	Student	Text
Curriculum	Curriculum	Student 1	Unit 1 Introduction
Curriculum	Curriculum	Student 1	1.1: Why Learn Programming for Music?
Curriculum	Curriculum	Student 1	1.2: Tools of the Trade: DAWs and APIs
Curriculum	Curriculum	Student 1	1.3: The EarSketch Workspace
Edit	Editing Code	Student 2	
Success	Compile	Student 2	Success
Success	Compile	Student 1	Success
Edit	Editing Code	Student 2	
Curriculum	Curriculum	Student 1	1.4: Running a Script
Curriculum	Curriculum	Student 1	1.5: Adding Comments
Curriculum	Curriculum	Student 1	1.6: The DAW in Detail

Out of the two curriculum usage methods we identified—targeted and undirected—the *Curriculum Browsing* state is characterized by the latter. In the pairs that have a high relative frequency of *Curriculum Browsing* → *Code Editing* transitions, typically one partner browses the curriculum while the other works in the code editor, and the contributions to the artifact are unequal. The excerpt previously referenced in Table 4-8 depicts **Student 2** working in the code editor while **Student 1** browses the curriculum, creating a transition from *Curriculum Browsing* → *Code Editing*. The negative relationship between the transition from *Curriculum Browsing* → *Code Editing* and *combined satisfaction* suggests that when one student is unsure in a way that causes this undirected curriculum usage, the student may require extra support.

4.4.3 Aesthetic Dialogue → Technical Dialogue: Negative Relationship with Combined Satisfaction

Previously we observed that the typical progression of a session begins in the *Social Dialogue* state, then pairs typically transition to the *Aesthetic Dialogue* state. Next, pairs move on to the *Technical Dialogue* state, which is followed by the code editing states (Figure 3-4). However, the regression model (Table 4-1) showed a negative relationship between the relative frequency of the transition from *Aesthetic Dialogue* → *Technical Dialogue* and *combined satisfaction*. To understand why the transition *Aesthetic Dialogue* → *Technical Dialogue* would have a negative relationship with *combined satisfaction*, we will review prior literature including cognitive load and the dynamic between learning and making, then ground this discussion in excerpts from the corpus.

Mastering basic coding skills already creates a large cognitive load on the student, and the addition of a creative component greatly increases the cognitive load [64, p.27]. Programming is considered to have a high intrinsic load, which is the difficulty inherent in a task [48, p.257]. Collaborative learning can help reduce this cognitive load by introducing a *collective working memory*, with communication being essential to its creation [93, 94]. In our corpus, as *Code Editing* and the debugging process increases, dialogue decreases, which can be seen in the ridgeline depicting the frequency of the hidden states over the session in Figure 3-4. The decrease in dialogue is more apparent when we examine the pairs with the lowest *combined satisfaction*. However, when we look at the pairs with the highest *combined satisfaction*, dialogue is relatively consistent throughout (Figure 4-3). Research indicates that during complex tasks, collaboration works as a scaffold for the knowledge acquisition process of each partner, and if it does not, it can become ineffective by causing too much extraneous load [94]. Seeing a drop in dialogue when coding and debugging begins is unsurprising because research has found that the most difficult programming issue for students is *finding bugs in their own programs* [101]. Some possible explanations for observing more consistent dialogue throughout the sessions of the pairs with the highest *combined satisfaction* are that these pairs did not have as much trouble with the

programming aspects, like debugging, or that the partnership was effective at reducing some of the cognitive load. In the next three paragraphs we will discuss excerpts from the corpus from a cognitive load perspective.

One difference we observed between the pairs with the highest and lowest *combined satisfaction* was in how the pairs approached the task. The pairs with the highest *combined satisfaction* tend to focus on their task requirements first, preferring to get simpler code elements in place before working out the details of the music. In the excerpt in Table 4-9 we can see just that—Student 1 making suggestions about the sounds they should use, but *Student 2* wants to focus on making the loop before playing with the sounds. Breaking the task down into smaller more manageable parts by focusing on the basic coding elements first may help reduce the cognitive load in two ways. First, this resembles the iterative refinement process where the program is built in chunks, includes testing and debugging before adding another level of complexity, and is one of the strategies that supports the debugging process [64, p.262]. Second, focusing on the aesthetic and technical concerns separately may keep students from being overburdened with the cognitive load of addressing both components simultaneously.

Table 4-9. Student collaborative dialogue excerpt presenting prioritization between technical and aesthetic considerations. Each state was determined automatically using the HMM presented in this work.

State	Action	Student	Text
Aesthetic	Proposal	Student 1	we should put like a synth or something to that effect
Aesthetic	Proposal	Student 1	add some like futuristic noises or airhorns or something
Aesthetic	Proposal	Student 2	we need to add the stuff to make it a loop first and then play with the sounds
Technical	Statement	Student 2	so we have the essential stuff done
Technical	Acknowledgment	Student 1	yeah
Technical	Editing Code	Student 2	
Technical	Directive	Student 1	im gonna put each sound into a variable to make it easier to read

The iterative refinement process is demonstrated by the pair featured in Table 4-10, where initially, *Student 2* suggests making a function with no arguments and then assigns the individual pieces of the function to each partner. By starting with no arguments in the function, the pair is

able to get the basic elements in place and working before increasing the complexity.

Additionally, this pair's dialogue was very consistent throughout the entire session with more of their *Technical Dialogue* occurring at the beginning of the session and more of their *Aesthetic Dialogue* occurring in the third quarter of the session.

Table 4-10. Student collaborative dialogue excerpt presenting initial planning and role assignment. Each state was determined automatically using the HMM presented in this work.

State	Action	Student	Text
Social	Social	Student 1	hey
Social	Social	Student 2	hey
Aesthetic	Proposal	Student 1	rock theme?
Aesthetic	Proposal Accept	Student 2	sounds good
Technical	Acknowledgment	Student 1	alright
Technical	Directive	Student 2	lets start with just making a function with no arguments first
Technical	Directive Accept	Student 1	k
Technical	Directive	Student 2	you make the fitMedia
Technical	Directive Accept	Student 1	ok
Technical	Directive	Student 2	i can do the loop
Technical	Directive Accept	Student 1	alright

In comparison, another pair begins their session by discussing the aesthetic parts of the project, assigning roles to each partner, but do not follow the iterative refinement process. The pair does not attempt to compile their code until 20 minutes after the session starts when they already have multiple functions written, which may have contributed to the difficulty they had trying to fix the syntax error they were receiving. The entirety of the pair's interactions, starting from about half of the way through the session, can be seen in the excerpt in Table 4-11, which includes their first attempt to compile during the session. Towards the end *Student 1* asks "whats wrong with our functions." The pair did not have a single successful compilation during their entire session.

Because adding an aesthetic component can increase the cognitive load of the programming project, focusing on the aesthetic and technical elements individually may help the pairs reduce their cognitive load. Novices in particular may face increased cognitive load from the technical portion of their programming tasks because they are still working to acquire necessary schemata to handle the technical portion of their projects [48, p.258]. This compartmentalization of the

Table 4-11. Student collaborative dialogue excerpt presenting error resolution attempts near session conclusion. Each state was determined automatically using the HMM presented in this work.

State	Action	Student	Text
Technical	Editing Code	Student 1	
Technical	Statement	Student 2	iv'e only found two sounds to use so far
Technical	Editing Code	Student 1	
Technical	Directive	Student 1	u can put them if you want
Technical	Editing Code	Both	
Technical	Confusion	Student 2	do i make another function
Technical	Statement	Student 1	yeah
Edit	Editing Code	Both	
Error	Compile	Student 2	SyntaxError
Edit	Editing Code	Both	
Error	Compile	Student 1	SyntaxError
Technical	Directive	Student 1	can you fix the mistake
Technical	Directive	Student 1	on line 20 and 21
Technical	Directive	Student 1	i mean 19 and 20
Edit	Editing Code	Student 2	...
Error	Compile	Student 2	SyntaxError
Error	Compile	Student 1	SyntaxError
Error	Confusion	Student 1	whats wrong with our functions
Edit	Editing Code	Student 1	
Error	Confusion	Student 2	i dont know whats wrong
Edit	Editing Code	Student 1	
Error	Compile	Student 1	SyntaxError

aesthetic and technical components can be seen in this utterance from a pair with a higher combined satisfaction: *“Now that we have a function we can call it multiple times but we can just change the sound.”* After their function is working, they can turn their attention to getting the song to sound the way they want. The benefits can be seen later in the session when they are working on the artifact aesthetic when one student says *“Oh I see the problem... When we try and call the function again the piano tries to start playing again on track 3, but its already playing”* to which their partner replies *“we need to add the thing that changes the track.”* While this still involves changes to the code, being able to compile their code lets them focus on how to change the sound without having to deal with coding errors. As noted previously, when students enter the

Compilation Error state, they tend to focus on resolving these errors instead of discussing the artifact aesthetics.

While this separation of *Aesthetic Dialogue* and *Technical Dialogue* seems to work for pairs that focus on the technical first, it may not work as well for pairs initially focusing on the aesthetics. In Table 4-12 we can see a pair listening to and deciding what sounds to use. One partner suggests making a loop, but the other partner says they should get all the instruments and then make it a song. When they do try to move onto making it a song towards the end of the session, **Student 2** says “*start finding other instruments, i will turn this part into a base.*” Afterwards it looks like **Student 1** finds a sound and says “*try it out now,*” but they do not get a chance to discuss it as the session ends a minute later. This pair seems to have gotten stuck on the aesthetics and were unable to progress to putting the song together. This evokes another critique of creative coding—that it can move the focus from learning to making [64, p.26], creating what has been referred to as the ‘virtuous’ and ‘vicious’ cycles of STEAM CS education [123]. The ‘virtuous’ cycle involves being motivated by the aesthetic part of the artifact to learn more advanced coding concepts to continue improving the artifact. The ‘vicious’ cycle happens when students learn just enough to be able to make the artifact and focus almost entirely on the artifact aesthetics and do not progress their computational learning. In the previous example, the students knew enough to be able to play the sounds they picked, but did not use custom functions, loops, or even vary the measures of the sound clips—all clips started in measure 1 and ended at 11—to turn the picked sounds into a song.

When students put too much focus on the aesthetic without having a solid foundation of the fundamentals of programming, students’ aesthetic ambitions can exceed their technical capabilities [64, p.27]. While some challenges can motivate students to put more effort into the activity, if they are unable to unresolved frustration can lead to boredom and disengagement [44]. One reason the higher rated pairs may benefit from focusing on the technical aspects first is that it may keep students’ aesthetic ambitions more realistic for their programming proficiency. This can keep the challenges students face while working on their artifact motivating instead of frustrating. One pair with lower *combined satisfaction* in the excerpt in Table 4-13, starts the session heavily

Table 4-12. Student collaborative dialogue excerpt presenting decision patterns in sound selection prior to code implementation. Each state was determined automatically using the HMM presented in this work.

State	Action	Student	Text
Success	Compile	Student 1	Success
Success	Pos Feedback	Student 1	woah thats neat
Aesthetic	Proposal	Student 2	do you want to use those drums
Aesthetic	Proposal Accept	Student 1	u can put them if you want
Aesthetic	Acknowledgment	Student 2	ok
Aesthetic	Proposal	Student 1	so like do you want to make a loop
Aesthetic	Proposal Reject	Student 2	not yet
Social	Social	Student 1	wait nevermind
Social	Social	Student 1	oh
Technical	Directive	Student 2	first lets just get all the instruments then we make it a song

focusing on the aesthetic details, even suggesting making a jingle for a brand. While they have some success early on, they begin running into more errors as they get into the final quarter of the session. Eventually *Student 2* says “*it looks fine. i don’t know why its not working. we should look back into unit one*” and begins looking through the curriculum. They are able to get some successful compiles by removing code, but they end the session with a *TypeError* after adding more code in. [125] found similar behavior when pairs became stuck during debugging.

Focusing on making sure the code works first can also ensure the pairs are able to work on the aesthetic portions of their song. This is because if the code will not compile, the students cannot hear their song. The excerpt referenced previously in Table 4-2 shows one student ask their partner if they want to make any changes to the song, and the other student says they need to listen to it first. They are unable to compile the code and thus unable to listen to the song they created. Then they spend the rest of their session debugging. Encountering technical errors when trying to work out the aesthetic aspects of the artifact can force students to focus on the error and potentially frustrate students that would prefer to work on the artifact aesthetics. In contrast, a pair previously referenced finished their function first and was then able to focus on the aesthetic aspects of their artifact like changing the track. This may be why the pairs that tended to get the basic code elements working first had higher *combined satisfaction*.

Table 4-13. Student collaborative dialogue excerpt presenting initial focus on aesthetic choices before technical implementation. Each state was determined automatically using the HMM presented in this work.

State	Action	Student	Text
Aesthetic	Passing Responsibility	Student 1	what kind of song should we do
Aesthetic	Passing Responsibility	Student 2	Do you want to use a song with a lot of beats or slower tempo?
Technical	Statement	Student 1	I normally use the same tempo
Technical	Confusion	Student 2	what type is that?
Technical	Statement	Student 1	the same 120
Aesthetic	Passing Responsibility	Student 2	okay what sounds do you want to use?
Aesthetic	Proposal	Student 2	we could start with some percussion to have a beat to start off
Aesthetic	Proposal Accept	Student 1	that sounds fine with me
Aesthetic	Proposal	Student 2	do you want some type of drum or piano?
Aesthetic	Proposal Accept	Student 1	drums sound good
Aesthetic	Passing Responsibility	Student 1	what's our theme tho
Aesthetic	Editing Code	Student 1	
Aesthetic	Confusion	Student 2	i don't know
Aesthetic	Proposal	Student 2	what about a jingle for a brand
Technical	Statement	Student 1	We'll figure that out later then

In this section, we discussed cognitive load and the dynamic between learning and making as possible reasons for the negative relationship between the transition from *Aesthetic Dialogue* → *Technical Dialogue*. Using excerpts from our corpus, we highlighted some of the differences between the pairs with the highest and lowest *combined satisfaction*, which include use of iterative refinement and focusing on task requirements first. We noted that the pairs with the highest *combined satisfaction* had more consistent dialogue throughout the session, including during the debugging phase when dialogue decreased considerably between other pairs. Some reasons for the higher levels of dialogue during the debugging phase could be that the pairs had less trouble with the programming aspects, like debugging, or that the partnership was effective at reducing some of the cognitive load. Additionally, focusing on task requirements first seemed to allow the pairs that did this to put more focus on the aesthetic elements in the later half of the session, whereas several of the pairs that focused heavily on aesthetics early tended to spent the later half of the session dealing with compile errors. If students are more motivated to discuss problems involving

the aesthetic elements rather than compile errors, it could explain some of the increased dialogue in the pairs with higher *combined satisfaction* since they tended to be able to focus on aesthetic problems in the later half of the session. Frustration with debugging and a desire to work on the aesthetic portions may also explain some of the virtuous/viscous cycle discussed previously.

4.5 Implications for Intelligent System Support

The findings of this research, while preliminary, provide insights into how co-creative processes are associated with *combined satisfaction*, and they illuminate potential opportunities around compilation errors, uncertainty, and the dynamic between learning and making to support collaboration during co-creative learning tasks. Some of these findings are specific to the co-creative process, while others apply to collaboration in general, and thus, have important implications for both co-creative agents and collaborative support systems. We will discuss each finding with its implications for both co-creative agents and collaborative support systems, but first we will define and briefly discuss the role of these two types of intelligent systems.

Co-creative agents fulfill the role of collaborator on open-ended creative tasks. Using agents as partners to human learners has demonstrated benefits including significantly higher levels of shared understanding, progress monitoring, and feedback [146]. Applications of co-creative agents range from improving collaborative skills, like encouraging “deep thinking” and initiative taking [72], to providing emotional support for elementary school students learning to code [124]. A co-creative agent can support human creativity through supporting the creation of music through code [180] and supporting collaborative ideation through sketches [106].

Collaborative support systems are systems where the primary role is to support collaboration between humans. Some of these systems perform functions like encouraging productive dialogue during collaborative learning [43, 177], while others help students regulate their collaboration by prompting students to use different reflection strategies [150].

4.5.1 Support for Error Resolution

Our findings indicate that the pairs with more interactions in the *Compilation Error* state tend to rate their partners lower than pairs with fewer interactions in the *Compilation Error* state.

We identify patterns that were common in the higher rated pairs that are frequently missing in the lower rated pairs: they tend to have more of their errors resolved by the end of the session, focusing on a single task together, explicitly stating when they are done and ready to move on, and work together to resolve uncertainty even if the coding error is already resolved. We can use these observations from more successful pairs to guide intelligent system support.

Because dialogue was such an effective tool for resolving errors, if co-creative agents detect an extended amount of time spent in the *Compilation Error* state, they could try to engage the student in discussion about the error and task. If the student is still unable to resolve the issue after talking through the issue, the co-creative agent could provide extra help resolving programming errors or offer examples of working code so that the student can remain engaged with the aesthetic aspects of the work.

For collaborative support systems, support could look like ensuring both partners are working on the same problem if there is an extended amount of time spent in the *Compilation Error* state. That could involve prompting partners to explicitly state what they believe the problem is and encouraging them come up with a plan for how they will fix it. Additionally, after the error is resolved, the system might have partners confirm with each other what the issue was and what fixed the error. This would give both students the opportunity to benefit from the learning experience.

4.5.2 Support for Help Seeking

The behavior of using curriculum documents in the way that characterizes the *Curriculum Browsing* state—where a student clicks through many different topics instead of seeking out a specific topic—can potentially identify when a student is having trouble getting the help they need. However, our findings show that lower *combined satisfaction* was only associated with the relative frequency of transitions from *Curriculum Browsing* → *Code Editing*, so this browsing behavior itself does not necessarily indicate issues in the collaboration. This transition was most common when one student would work in the code editor while their partner browsed the

curriculum. In these situations, contributions to the artifact were very unequal and conversation was limited, and in many cases the editing student ignored questions from the browsing student.

While a co-creative agent would not likely be designed to work on an artifact without input and contributions from the human partner, there are still things to be learned from this finding. If an agent notices this type of curriculum browsing behavior, they could try to find out what the student is having trouble with and potentially offer resources, ideas, or encouragement. For example, if the problem is related to the aesthetic aspects, like how to decide on a genre, the agent can help the student brainstorm or give examples of different sounds. At a minimum, co-creative agents should treat this curriculum browsing behavior as a indicator to check in with the student.

Alternatively, when collaborative support systems notice this behavior, it could indicate problems with collaboration between the partners. Systems can encourage the browsing partner to communicate any uncertainty they have with their partner. Systems could also monitor chat and artifact contributions. If the browsing partner has tried to communicate, but the editing partner has not responded, the system could give a nudge or reminder to draw the attention of the editing partner to the browsing partner's message. If the browsing partner has not made many contributions to the artifact, it could encourage the pair to set specific tasks for each partner. In extreme cases, like where one partner ignores the other and possibly even removes the browsing partner's contributions, an alert to the teacher about the issue may be warranted so they can check in on the pair.

4.5.3 Scaffolding Aesthetic and Technical Elements

Prior research has shown that when students talk through what they want to make first, it prevents students from putting too much emphasis on project requirements and not enough on personal relevance [64, p.23]. Additionally, many students can be more motivated by the personally relevant aspects of the project, like creating music, than the programming itself [111]. Furthermore, when many of the aesthetic decisions of a project are left to the end, they can require major changes to the technical aspects. In EarSketch, this may look like changing or adding new code elements to add a drum fill effect to a song. This can potentially lead to sticking with an

aesthetic decision that is not preferred to avoid rewriting the program or being unable to finish the changes required if students run out of time to work on their artifact. However, despite the benefits of including aesthetic elements in programming, our findings suggest that students could likely use more support during the co-creative process in two main areas—reducing cognitive load and scaffolding the aesthetic process.

For agents acting as co-creative partners, students can benefit from the knowledge scaffolding provided by collaboration by having a partner that can provide the right knowledge at the appropriate time [94]. For example, providing relevant worked or partial worked examples [202, 117] and encouraging good coding and debugging habits as they move through the project are ways that can reduce the cognitive load of the technical portion of the task and allow students to spend more time on creating. To help with the aesthetic elements, co-creative agents could use details from the artifact and make suggestions or give examples for moving what the student already has done towards a completed version of the artifact. For example, if students are just inserting sounds without progressing to a song-like structure, agents could suggest a custom function to make a chorus with sounds already present in the code editor or even provide one. This could give students a clear direction for moving forward on completing their artifact.

Collaboration support systems can identify when pairs encounter issues through dialogue features [60], such as dialogue decreasing when errors increase, and encourage partners to communicate. They can even help guide the conversation when pairs are not having productive conversation [43]. These systems can also encourage pairs to set explicit goals or tasks—like using a custom function—and help them track which ones they’ve accomplished. Similar to the co-creative agent, these systems could also suggest more advanced code structures if the pair seems not to be making progress towards completing the artifact.

4.6 Key Takeaways for My Dissertation Research

This chapter presents an investigation of the relationship between the co-creative states coupled with their transitions and partner satisfaction. The results indicate the importance of regular dialogue in successful co-creative partnerships, especially when navigating challenges.

More satisfied pairs maintained consistent dialogue throughout their sessions, while less satisfied pairs' dialogue tended to decrease after they transitioned to working in the code editor (illustrated in Figure 4-3).

While my prior research investigates the structure and outcomes of human-human co-creative processes, the rapid advancements in AI, particularly in conversational models like GPT-4, necessitate a fresh investigation into human-AI co-creativity. This work informs the development of a co-creative AI (CAI) partner for the EarSketch environment.

CHAPTER 5

DEVELOPING A CO-CREATIVE AI PARTNER FOR CREATIVE COMPUTING

This chapter presents the development and evaluation of a Co-creative AI (CAI) system designed to support students learning to code while creating music in EarSketch.¹ Building on our understanding of human-human co-creativity (Chapter 3) and its relationship with partner satisfaction (Chapter 4), we developed and refined CAI through five iterations of design and evaluation in real classrooms. This work represents a crucial step in understanding how AI can support human creativity in educational contexts, providing insights that directly inform the design of my dissertation study examining AI’s role in narrative writing (Chapter 6).

5.1 Background and Motivation

The expanding role of AI in education presents unique challenges when designing conversational agents that support users in creative, open-ended learning environments. As computers have evolved from tools that merely support creative work to active participants in the creative process, new forms of human-computer collaboration have emerged. This evolution has given rise to human-computer co-creativity, enabling humans and computers to jointly contribute to creative artifact and idea production [35].

The design of creative, open-ended learning environments is informed by constructionism, which posits that learners enhance their knowledge through the creation of external artifacts [69]. In Computer Science (CS) education, educators have used constructionism alongside pedagogies such as inquiry-based education and design science research to support computational thinking [41]. Modern CS education at the K-12 level embraces various approaches that cultivate computing skills and knowledge through constructionist platforms for making and sharing expressive artifacts, including music [102], games [175], animations [31], and robotics [197].

EarSketch exemplifies this approach as an expressive CS learning environment that engages predominantly middle and high school learners in creating computational music with JavaScript or Python code. The platform has demonstrated success in yielding learning gains and positive changes in learner attitudes towards CS, especially among underrepresented students across

¹ EarSketch is available online at <https://ears sketch.gatech.edu>

gender and ethnicity [112]. However, like other CS environments [66], students often understand just enough to create basic artifacts but rarely develop deeper computational thinking skills such as abstraction, logic, and iteration.

While recent CS education research has focused on innovative approaches used by constructionist learning platforms, research on AI-based assistance for learning programming remains focused on more traditional programming learning environments. Traditional approaches to AI assistance in CS education, like intelligent tutoring systems [8] and learning companions [73], are designed for well-defined problem spaces. These interaction models break down in open-ended creative contexts where there is no “correct” solution or output. In EarSketch, for example, a typical assignment might be to “make a song with 64 measures and a verse-chorus structure” rather than “sort this list of numbers.” In this context, an AI agent must help students discover their goals, scaffold their actions, and troubleshoot their problems while supporting both technical and creative aspects of the task.

To address this gap, we developed CAI—a co-creative learning companion for EarSketch that aims to support both the technical and creative aspects of computational music creation. Through five iterations of design and classroom evaluation, we investigated the following research question: *What challenges and opportunities emerge when designing and evaluating a co-creative AI for an expressive CS learning environment?*

5.2 System Description

5.2.1 EarSketch Overview

The EarSketch web client contains a Code Editor, Digital Audio Workstation (DAW), Content Manager, and Curriculum (see Figure 5-1). Users create Python or JavaScript projects and edit them in the Code Editor, using EarSketch API functions such as `fitMedia()` and `makeBeat()` to arrange sound samples onto audio tracks for a given duration and apply audio effects. Once they compile the code, the DAW displays the result and the user can play, pause, loop, and solo or mute individual tracks to hear their project.

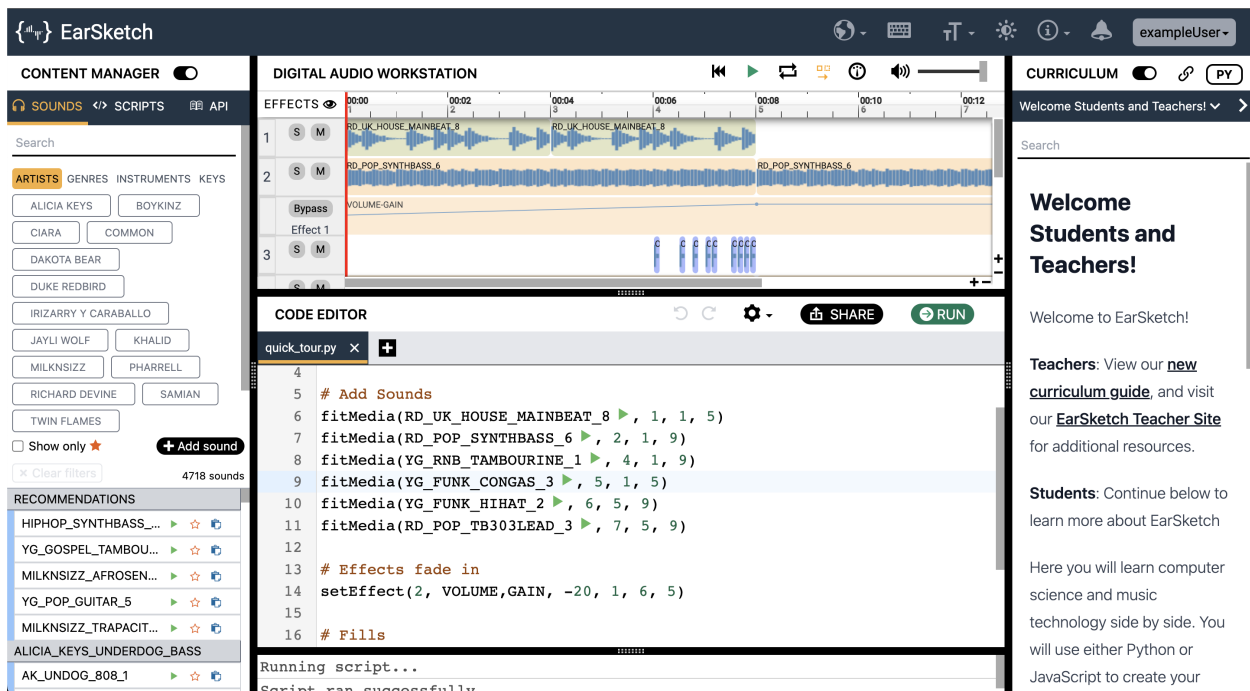


Figure 5-1. A screenshot of the EarSketch interface, with Sound Browser (left), Digital Audio Workstation (top), Code Editor (bottom), and Curriculum (right).

Users can search for sounds to use, manage the multiple projects they have created, and search for EarSketch API functions and documentation in the Content Manager. The Sound Browser contains a text search as well as genre, artist, instrument, and key signature filters. Additionally, a recommendation system uses the contents of an active project to present complementary sounds to a user based on audio analysis as well as combinations of sounds used by others [165, 163, 166].

The Curriculum pane on the right side of the interface contains lessons and example code for various coding and music concepts such as loops, function definitions, musical form, debugging, and peer review. These lessons have been designed to be integrated with a full computer science curriculum, as well as the “Hour of Code” module for a single-session introduction to creative computing. Additionally, the EarSketch site has been the stage for multiple Your Voice Is Power competitions in partnership with Amazon Future Engineer² designed to increase student involvement in CS education with themes of social change and equity.

² <https://www.amazonfutureengineer.com/yourvoiceispower>

5.2.2 CAI Overview

The CAI (Co-creative AI) interface resides on the right side of an experimental version of the EarSketch web client (Figure 5-2), replacing a user’s view of the Curriculum while it is active. CAI initializes a separate conversation related to each script a user has opened, displayed to the user with the heading “Talk to CAI about [project name]”. A toggle button can be used to switch views between CAI and the curriculum in the header at the top of the EarSketch window. Conversations between a user and CAI are depicted as text message-style bubbles originating from opposite directions, owing to CAI’s modeling after human-human collaboration. Users can send messages to CAI through dialogue buttons, which are generated in response to the user’s place in CAI’s internal dialogue tree [181] or by triggering events such as compiling code with an error. CAI sends messages to a user in response to user actions, such as running a script, as well as instantaneously responding to a user’s input messages.

5.2.3 CAI Subsystems

CAI consists of four main modules (Figure 5-3). The **Analysis Module** includes a Complexity Calculator that evaluates mastery of EarSketch concepts by parsing the abstract syntax tree of user projects [164]. It also generates Sound Profiles representing musical structure through section analysis. The **Dialogue Module** manages user conversations using Sound Profile and Complexity Calculator outputs to inform responses from a dialogue tree with over 150 nodes. The **Error Handling Module** generates explanations for common Python and JavaScript errors. The **Suggestion Module** recommends code and musical changes based on weighted random selections influenced by user actions.

5.2.4 Integration with EarSketch

CAI uses a React³ interface, as does the rest of the EarSketch web client. It uses redux⁴ state management to update the interface as state variables are stored and functions are resolved, such as with the generation of dialogue. This allows CAI to work without interrupting other processes

3 <https://react.dev/>

4 <https://redux.js.org/>

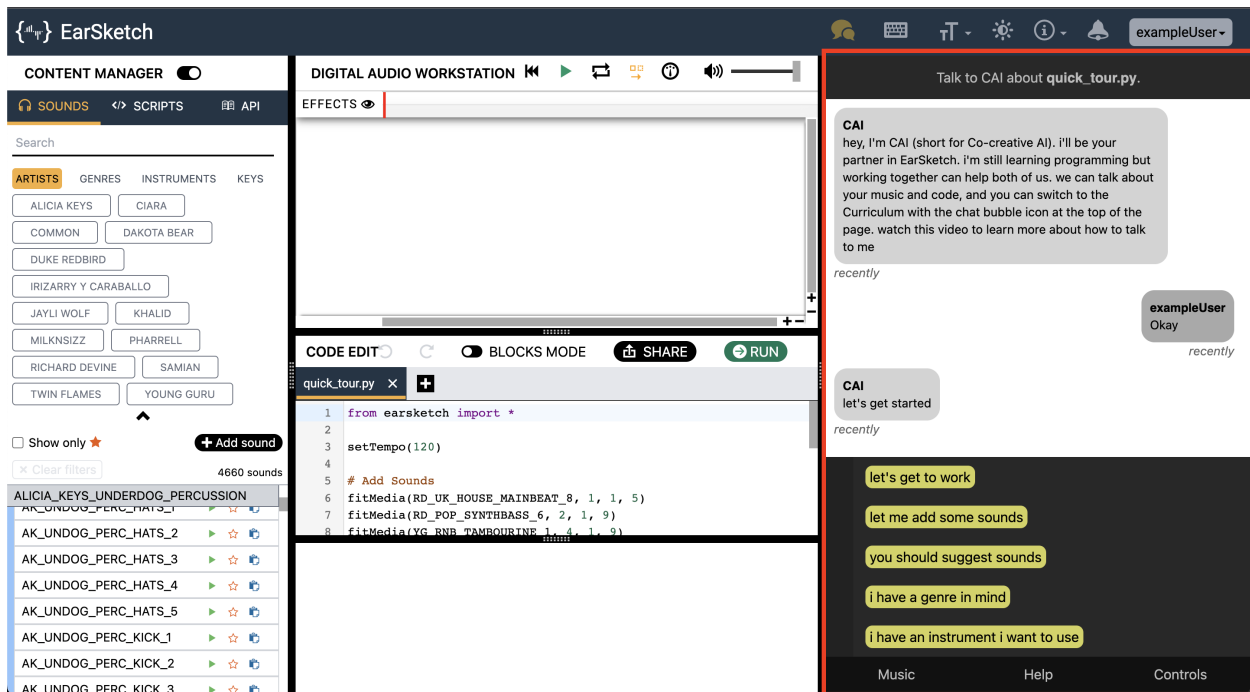


Figure 5-2. The EarSketch interface with CAI enabled (right).

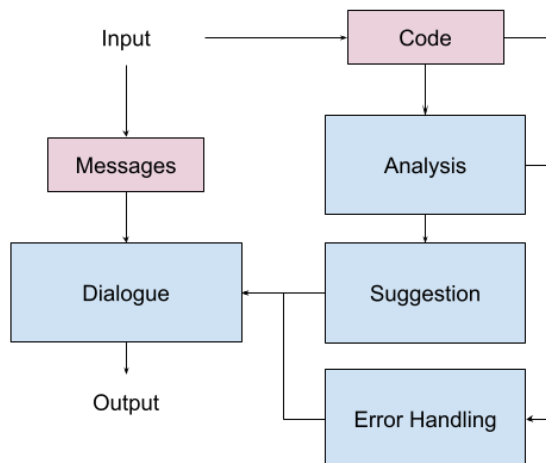


Figure 5-3. A simplified architecture diagram showing how the modules of CAI interact with user input code and messages to generate dialogue output.

vital to the EarSketch user experience, such as fast retrieval of audio samples from a server or responsive Digital Audio Workstation controls.

Additionally, since CAI is not intended to be used in the official version of EarSketch used in production, a series of environment flags were designed to allow for rapid deployment of specific test versions of CAI to be used in studies: the presence of the CAI interface, user-to-user chat, and the collection of keystroke-level user interaction data are all able to be configured and uploaded to various URLs for study-specific EarSketch versions. The data logs, including connections from individual student usernames and the messages sent to and from CAI, can be monitored in real time by accessing a MySQL database specific to these special deployments.

For analysis of collaborative behaviors, we collected students' interaction data, including keystrokes from the code editor, chat messages, and interface actions such as code compilation and curriculum access. This data was stored in a database with timestamps and anonymized project and sender IDs.

5.3 Human-Human Studies as Foundation for CAI

The development of CAI was directly informed by our understanding of human-human co-creativity gained through the studies presented in Chapters 3 and 4. Our analysis revealed several critical patterns in successful peer collaboration that shaped CAI's design and functionality. These patterns, detailed below, encompass the role of formative project planning, the balance between technical and creative aspects, and approaches to error resolution.

5.3.1 Importance of Formative Project Planning Discussions

Observing how student pairs typically moved through these states revealed that the co-creative process often began with casual social interaction and then transitioned into a project planning phase. Pairs typically began the project planning phase by discussing their creative vision (e.g., *“let's do dubstep cause its fire”* or *“what about a jingle for a brand”*) followed by discussing project requirements and initial implementation approaches (e.g., *“i think we need to make a function and call it twice idk how though”* or *“we should have variables for music 1 and 2”*). A particularly concise example of this process is shown in the excerpt from one pair's

dialogue in Table 5-1, with the pairs starting with brief greetings, confirming a genre, and then confirming a technical plan and assigning tasks.

Table 5-1. Student collaborative dialogue excerpt showing typical progression through the formative project planning stages, from social interaction to aesthetic decisions and technical planning.

State	Student	Text
Social	Student 1	hey
Social	Student 2	hey
Aesthetic	Student 1	rock theme?
Aesthetic	Student 2	sounds good
Technical	Student 2	lets start with just making a function with no arguments first
Technical	Student 1	k
Technical	Student 2	you make the fitMedia
Technical	Student 2	i can do the loop

5.3.2 Balancing Technical and Aesthetic Components

From our analysis of the students' dialogue and system logs and their relationship with the partner satisfaction survey, we noted the difference in how the pairs balanced the technical and aesthetic aspects of their project. Pairs reporting higher satisfaction scores generally approached their project iteratively, establishing basic technical foundations while maintaining flexibility in their aesthetic choices. These pairs would typically establish a general creative direction during the planning phase, such as choosing a genre or style, but avoided detailed aesthetic decisions until they had basic functional coding structures, like those required by the activity instructions, in place.

5.3.3 Proactive Error Support

Our analysis showed that pairs with higher partner satisfaction demonstrated distinct patterns in handling technical challenges. These pairs quickly addressed compilation errors and maintained active communication about their planned actions. When encountering errors, these pairs would follow a typical pattern: identifying the issue (e.g. *"line 1 had an error"* or *"it says its missing a ; before statement"*), collaboratively working toward a solution (e.g., after receiving an error S1: *"idk if it has to do with the numbers or not"* S2 responds: *"the numbers are right i just think its something with a comma"*), and then confirming the their understanding of the

resolution before proceeding (e.g., after a successful compilation, clarifying “*It was called I just had to add the fitMedia within the function*”).

These insights directly informed the design of CAI’s core functionalities and interaction patterns. For a complete analysis of the co-creative states and their relationship with partner satisfaction, see Chapters 3 and 4.

5.4 First CAI Prototype

The initial deployment of the CAI prototype aimed to closely emulate the supportive, interactive dynamics of human-human collaboration. Building on insights from our study on human-human collaboration and broader Human-Computer Interaction (HCI) literature, we designed the first CAI prototype to proactively assist students with compilation errors, support both the technical and aesthetic aspects of their projects, and facilitate early project planning discussions.

5.4.1 System Design

As shown in Figure 5-4, we implemented these capabilities through a menu-based chat interface that guides students through common collaboration patterns. The interface serves as the central integration point, where the Analysis Module’s ongoing evaluation of both code complexity⁵ and musical structure informs the Dialogue System’s interactions. For example, when a student encounters a compilation error, the Analysis Module evaluates both the specific error and the overall project context, allowing the Dialogue System to provide targeted assistance that considers the student’s demonstrated knowledge level and project goals.

The Analysis Module implements this contextual awareness through two main components. The Code Complexity Calculator maintains a dictionary of curriculum concept usage, tracking elements like function definitions, control structures, and API usage patterns. The Sound Profile generator creates a structured representation of the project’s musical elements, mapping sounds and effects across measures and tracking musical structure. These analyses enable CAI to provide suggestions that are both technically appropriate and musically relevant.

5 The term “Code Complexity” as used here refers to the depth and breadth of EarSketch curriculum concepts applied within a user’s project, rather than traditional software engineering complexity metrics.

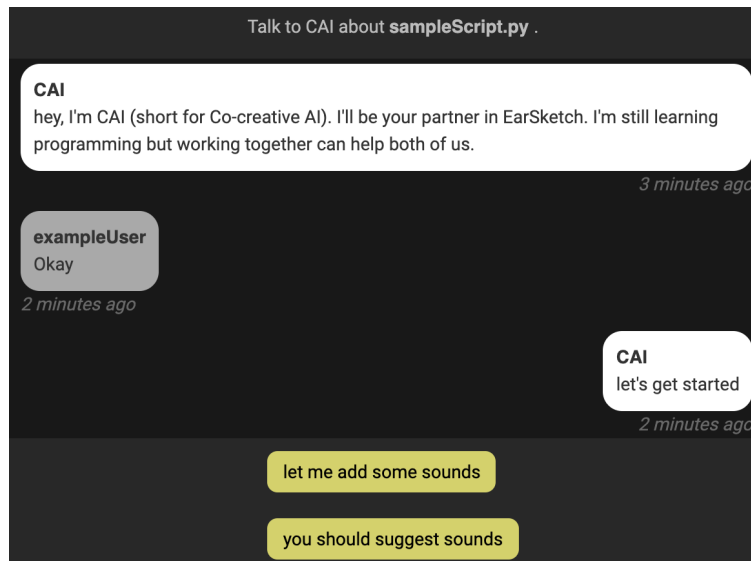


Figure 5-4. Screenshot of the first CAI interface illustrating the initial greeting from CAI and the subsequent user options to add sounds or request sound suggestions [181].

The Dialogue System uses these analyses to drive a menu-based interaction model that mirrors natural collaboration patterns. Starting with social interaction and project planning, the system progressively introduces technical and creative support options based on the student's activities and project state.

5.4.2 Study Design

To evaluate the CAI system's effectiveness in supporting students' music and coding activities and to identify opportunities for enhancements, we conducted a pilot study in February and March of 2021 across three high schools in the Southeastern United States [181]. Out of 56 students who agreed to participate, we gathered complete data sets from 42 participants. The demographic breakdown included: 55% White, 18% Asian, 18% Black, and 7% Hispanic/Latino, with 82% male and 18% female participants, ages ranging from 14 to 17. The majority of participants (64%) were in the 9th grade, with 20% in 10th, 14% in 11th, and 2% in 12th grades. All students had prior experience with EarSketch through their class curricula.

We adapted the study to fit both in-person and virtual learning environments due to the COVID-19 pandemic constraints. Students completed a coding assignment, accessing a special version of EarSketch that integrated the CAI interface, and completed pre- and post-surveys

focused on students' perceptions of CAI as a co-creative partner on a 4-point Likert Scale (Table 5-2). We also integrated open-ended questions in the post-survey that asked why they wanted to work with CAI again (if they answered that they would) and to identify the most valuable suggestion CAI made to enrich our understanding of the CAI's impact. Additionally, we captured data on students' interactions with CAI through enhanced data collection tools within CAI, providing valuable context for analyzing the survey results and assessing the system's impact.

Table 5-2. Questionnaire items measuring student perceptions of CAI as a co-creative partner, organized by assessment category.

Category	Question
Cognitive Support	CAI helped me do something I would not have done otherwise. CAI helped me make a better program. CAI helped me make a better song.
Interaction Quality	CAI made valuable suggestions. CAI said useful things to me. (open-ended) What was the most valuable suggestion CAI made?
Concentration/Flow	CAI frustrated me. CAI interrupted my work. CAI suggested things at the right times. I paid attention to CAI.
Overall Satisfaction	I enjoyed my interaction with CAI. I would like to work with CAI again. (open-ended) Why would you like to work with CAI again?
CAI's Competency	CAI's ability to help with music production (e.g. effects, changing volume) is... CAI's ability to help with my music (recommend a sound or musical form) is... CAI's ability to help write/debug code is...
Psychosocial Support	I felt encouraged by CAI. I was more motivated working with CAI than I would have been on my own. (open-ended) What's one thing CAI did not do that you would like it to do?

5.4.3 Findings

5.4.3.1 Dialogue system navigation

The CAI interface, positioned on the right-hand side of the browser window, shares a space with the curriculum panel. This setup requires the CAI interface to replace the curriculum panel when active, requiring students to toggle between interacting with CAI and accessing curriculum content. While this design choice facilitates a cleaner workspace, navigating between CAI's guidance and the curriculum can be cumbersome, frustrating, or overwhelming for users to leave the chat window to search through the curriculum for information when CAI suggests a curriculum resource. Additionally, if students navigate away from the CAI chat window, the interface lacks any indication of CAI sending a new message. These observations highlight the need for a more fluid interaction between CAI's interactive support and the static curriculum content.

In the post-surveys, students expressed appreciation for CAI's sound sample suggestions, highlighting CAI's ability to simplify navigating EarSketch's large sample library. This feedback highlighted CAI's role in making the music selection process more efficient, allowing students to find appropriate sounds without manually searching through the thousands of options available.

5.4.3.2 Limited scope of advice

Additionally, our review of interaction logs revealed that students tended to request sound sample suggestions more often than they sought advice on code or musical structure. This trend might reflect a greater need for assistance in selecting sound samples over coding or structuring music. However, it also points to potential areas for improvement. First, the frequent sound sample requests could indicate that students are trying multiple times to find a sample that meets specific criteria. Second, the fewer requests for code and musical structure guidance may be a result of the limited number of code and musical structure suggestions available [181].

These findings from the first prototype study highlighted key areas for improvement: the need for better integration between CAI and curriculum content, more sophisticated suggestion capabilities, and improved navigation. These insights directly informed the development of the

second CAI prototype, which focused on addressing these limitations while maintaining the successful aspects of the initial design.

5.5 Second CAI Prototype

Based on insights from the initial prototype study, we enhanced CAI to address three key challenges: the difficulty of switching between interfaces, the relevance of suggestions, and the limited variety of support offered. These improvements focused on creating a more integrated and personalized experience.

The curriculum integration addressed navigation challenges through two key mechanisms. First, by embedding relevant curriculum links within dialogue responses, students could quickly access learning materials contextually related to CAI's suggestions. Second, a flashing chat button was introduced to alert students to new messages from CAI, facilitating a smoother transition back to the chat interface.

We introduced two distinct enhancements to the suggestion system based on observed student preferences for music sample suggestions. First, we added filtering capabilities that allow students to limit sound suggestions based on specific criteria (genre and instrument) and location (sections or measure numbers of a song). This granular control enabled more targeted and relevant music sample suggestions, making it easier for students to find sounds that fit their specific needs.

Second, we developed a broader project requirements system where students could specify high-level goals for their project, including musical form, genre, instruments, and coding structures. Rather than just filtering existing suggestions, this system actively generated new suggestions to help students meet their stated goals. For example, if a student defined "HIP-HOP" as their genre alongside an "ABA" form, and their project lacked distinct sections, CAI would specifically suggest adding Hip-Hop samples that could help create the contrasting sections needed for the ABA form.

5.5.1 Study Design

We evaluated this revised CAI prototype in a follow-up study conducted in May 2021 involving students from two high schools in two states. This study aimed to investigate students'

interactions with the expanded functionality of the second CAI prototype and to assess the impact of the new features on the students' experience. A total of 21 students participated in the study, and 17 provided demographic data. Among those who shared, 59% identified as White, 18% as Hispanic/Latino, 12% as Asian, 6% as Black, and 6% indicated a background not listed in the survey options, with 23% female, 59% male, and 18% not identifying as either gender. The students' ages ranged from 14 to 17, with a majority of the students in the 9th grade (70%) and the remaining students in the 10th (12%) and 11th (18%) grades. Participants engaged in a coding activity using the second CAI prototype integrated within the EarSketch environment. The study utilized the same survey instruments as the initial prototype evaluation, focusing on students' perceptions of CAI as a co-creative partner.

5.5.2 Findings

Analysis of student feedback revealed broad appreciation for CAI's capabilities, particularly its music sample suggestions and error assistance. Students valued CAI's ability to streamline the song creation process, with one noting that "The sound suggestions there are so many options so the CAI narrowing it down made making the song so much easier." Others highlighted CAI's debugging support and its guidance during project initiation, with one student specifically appreciating CAI's help with "How to start because I have no idea." When asked about working with CAI again, students responded positively, elaborating on several key aspects of their experience detailed below.

5.5.2.1 Support and encouragement

Students felt supported and encouraged by CAI, with responses highlighting its helpfulness in generating ideas and guiding the music creation process. For instance, one student mentioned, "It was very helpful and brought up good ideas," while another said, "CAI was very helpful and encouraging to keep working on my project." Another student said, "It made the coding process a lot less stressful and really enjoyable."

5.5.2.2 Exploration and creativity

Students' feedback highlighted CAI's role in enhancing their creative process by introducing them to new possibilities and ideas. One mentioned, "It was something new and fun. It helped me create a new piece of music that I wouldn't have been able to create before," highlighting CAI's role in making their music project unique and enjoyable. Another said CAI helped "organize and recommend relevant things that I would not have thought of," showing how it guided them to ideas and sounds they hadn't considered, enriching their music.

5.5.2.3 Potential for growth

Some students acknowledged the potential in CAI despite recognizing its current limitations, expressing a desire to see how CAI evolves. One student expressed, "I feel that CAI has great potential and also makes the process of finding samples easier to do. Despite CAI sometimes recommending odd parts that didn't really fit at times, overall, the experience was good, and I would love to see how CAI continues to grow." Another shared their hope for CAI's improvement, saying, "I would like to work with CAI again because I see the potential in the program. It wasn't unhelpful. It simply needs work on making more useful suggestions."

5.5.2.4 Explanations and guidance

Although the responses were generally positive, students also provided constructive feedback on areas where CAI could be enhanced to better meet their creative and learning needs. Many students expressed a desire for more detailed and relevant suggestions from CAI, particularly in the context of debugging and creative decision-making. Some wished for CAI to offer clearer explanations on how and why certain additions or changes should be made to their projects. For example, one student wished for CAI to "Explain what it would want you to add better, like how and why you should add it," indicating a need for deeper insight into the reasoning behind CAI's suggestions. Additionally, there was feedback highlighting the need for CAI to provide more specific guidance on implementing complex programming concepts, such as adding "randomness" to a piece, with one student noting, "I wished that CAI would explain how to implement that code to help better complete the music."

5.5.2.5 Desire for open-ended dialogue

Students also expressed a desire to engage in more open-ended dialogues with CAI, allowing for a broader range of queries and interactions. This feedback suggests a need for CAI to support a more flexible interaction model, where students can ask questions outside the predefined menu options and receive responses that are tailored to their specific needs and contexts. For instance, a student expressed a wish to “be able to ask questions to CAI instead of what is in the suggested questions,” while others focused on limited menu options, stating that CAI “did not provide responses that I would have liked to say to the bot at certain points” while another wanted CAI to “Have more answer boxes.” Others expressed frustration with “repeat” suggestions by CAI that used a consistent template to suggest sounds and code changes statically. These responses highlight a desire for a more interactive and responsive CAI experience.

These findings, particularly the desire for more open-ended dialogue, led to our next phase of development: a Wizard of Oz study to explore more flexible interaction patterns.

5.6 Wizard of Oz Study

Following the second CAI prototype study, student feedback indicated a desire for open-ended communication with CAI rather than the existing menu-based system.⁶ To investigate this interaction method without fully developing natural language-based CAI, we conducted a Wizard of Oz study [32] using the functional CAI prototype alongside human intervention to simulate open-ended dialogue with a more advanced conversational agent.

5.6.1 Study Design

In the Wizard of Oz study, conducted during Fall 2021 and Spring 2022, students collaborated with researchers who pretended to be CAI. The CAI system was fully operational on the student’s web client, but instead of sending its messages directly to the student, it generated all possible messages for a given student prompt and sent them over a web socket to the researcher. The researcher could then select one of these options as an interface button or type their own

⁶ It is worth noting that this feedback was gathered prior to the widespread awareness and availability of advanced large language models (LLMs) among students. As such, their expectations for open-ended communication with AI systems were likely different from what they might be today, given the increasing exposure to LLMs in recent years.

message. Researcher “wizards” could also select from templates CAI uses to fill messages using keyword commands in the text input box. For example, the following command “sounds_good_suggest_sound” outputs

“sounds good. wanna try [sound_rec] next?”

with CAI generating appropriate sound recommendations in the text unless the wizard fills them in themselves. Figure 5-5 illustrates the flow of information between a user, CAI, and a Wizard during the Wizard of Oz study. Data logs were created for when wizards used CAI utterances, selected templates, or created their own.

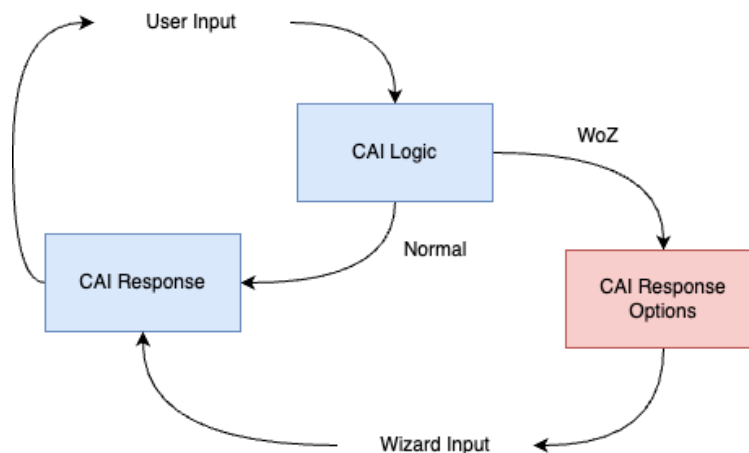


Figure 5-5. Flow of information in the Wizard of Oz study. The “Normal” path shows the standard flow, where user input is processed by the CAI logic to generate a response. The “WoZ” path shows the modified flow, where the CAI logic generates potential responses, which are sent to a human ‘wizard’ to select or provide a custom response.

The Wizard of Oz study took place over three sessions in high school classrooms, lasting for a full class period, each with a total of 36 participants. Following the study, the students completed the same survey instruments as the previous prototype evaluations.

5.6.2 Findings

Post-survey feedback revealed a few key issues with the wizard-operated version of CAI. Students reported frustrations with slow response times from the wizards, as well as inaccurate sound recommendations and a lack of debugging support. While students reported “repeat”

suggestions less often with the wizard-operated version of CAI compared to the menu-based CAI, they expressed a lack of starting suggestions with the wizard version. Additionally, we found that, despite having free textual input, students hesitated to initiate conversations about topics like debugging unless they repeatedly faced errors. *This observation suggests that while students are interested in the option to communicate freely with CAI, they may not always know how to ask for help or what kind of assistance they can request.* Based on these observations, we aimed to further develop CAI in terms of balancing flexibility and structure and providing onboarding to users.

5.6.2.1 Flexibility vs. structure

The decision to use a menu-based interface over natural language for CAI was influenced by the trade-offs between the two interaction styles. While a menu-based approach offers structured interactions, it can be challenging to craft code suggestions for different activities without limiting the system’s flexibility and usefulness. The predetermined options in the menu may not cover all possible student inquiries or creative directions, potentially hindering the co-creative process. On the other hand, a natural language interface might offer more flexibility and open-ended interactions, but ensuring the accuracy and relevance of code suggestions in real time can be a significant challenge. Natural language processing (NLP) also introduces complexity and can misinterpret user inputs, leading to incorrect or unhelpful suggestions [77, 54, 110].

While recent advancements in large language models have significantly improved NLP capabilities, at the time of this study, these technologies were not yet widely available or sufficiently reliable for our specific use case. As NLP technologies continue to advance, the specific challenges may change, but striking a balance between structure and flexibility remains an important consideration in the design of co-creative agents.

5.6.2.2 Onboarding and trust

Trust is an important factor in the effectiveness of a co-creative learning companion. It is defined as “the attitude that an agent will help achieve an individual’s goals in a situation characterized by uncertainty and vulnerability” [104] and covers the core aspects of trust [186]. In our Wizard of Oz experiments, we observed that learners had difficulty understanding CAI’s

abilities and limitations when trying to use it, which likely contributed to their lack of engagement with the system. Although learners were often interested in CAI, they wanted to better understand its capabilities and how to use them without an obvious way to do so.

These insights from the Wizard of Oz study directly informed the development of our third CAI prototype, which aimed to balance structured support with more flexible interaction options.

5.7 Third CAI Prototype

5.7.1 System Design

The Wizard of Oz study revealed important trade-offs between menu-based and natural language interaction approaches: while menu-based systems provide helpful structure, especially for novices, they can be limiting in their predefined nature. Conversely, natural language systems offer more flexibility but can be challenging for novice users who lack the domain knowledge to effectively discuss specific capabilities. Building on these insights and following established guidelines for AI capability invocation [5], we implemented significant architectural changes to achieve this balance.

The new menu system organized CAI’s capabilities into three main categories, informed by student survey responses from earlier studies (Figure 5-6). The “Music” menu provided direct access to sound recommendations filtered by genre or instrument. The “Help” menu offered branching explanations and examples for coding structures and EarSketch API functions. The “Controls” menu facilitated access to interface features like script history and curriculum navigation. This organization made CAI’s full range of capabilities more discoverable while maintaining structured access.

To improve system approachability, we implemented an onboarding system centered around an introductory video explaining CAI’s capabilities and interface navigation. The video aimed to build user trust and competence, with usage tracking integrated into server logs to understand its impact on subsequent user interactions.

We also enhanced three core components based on student feedback. The Suggestion Module now employs weighted random decision-making to generate recommendations across

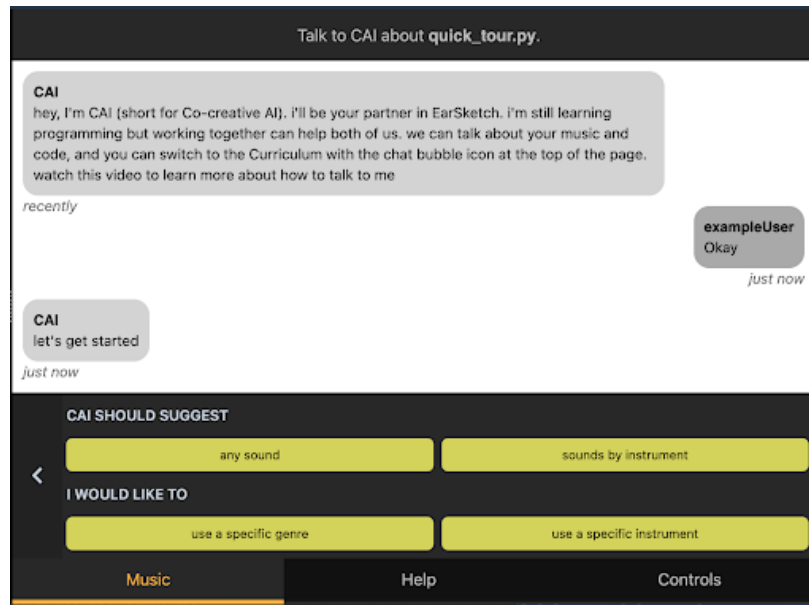


Figure 5-6. The CAI interface with the Music menu opened.

three categories: new coding structures, advancement of existing structures, and aesthetic/musical changes. Weights are dynamically adjusted based on user actions and suggestion history. The Error Handling Module was expanded to provide context-aware error solution suggestions by analyzing both raw code text and console error reports. Additionally, the Sound Recommendation system now incorporates musical key signature and beat complexity analysis [166], enabling suggestions based on musical compatibility and contrast while providing explanations for recommendations (e.g., users see sounds labeled with identical or complementary key signatures to those used in their projects).

5.7.2 Spring 2023: Classroom Studies

In Spring 2023, we conducted a study comparing students using CAI to complete a coding task against a control group using the regular EarSketch. This study took place in 18 sessions in high school classrooms, with data collected from 292 students. Students were given a coding task divided between two test versions of EarSketch: one with CAI enabled and one with the regular EarSketch interface but with data logging enabled. As with the previous studies, student interactions were monitored via live database analysis to ensure proper version usage as well as assignment completion. We expanded the post-study survey to include Intrinsic Motivation

Inventory [37] items based on the student’s enjoyment, perceived competence, effort, frustration, and value. Students were also asked if they had watched the introductory video embedded into the CAI interface.

5.7.3 Findings

Analysis of the system logs revealed differences in the interaction patterns between the CAI and non-CAI groups across various factors, including error handling in code, creating musical artifacts, and using prior work, such as previous scripts or projects. However, we noted that time spent interacting with CAI likely diverted attention from the other activities, highlighting an opportunity cost: introducing a learning companion may, at least initially, divert time spent on other aspects of the activity.

5.7.3.1 Adoption time

The lack of an accurate model of beginning CAI users contributed to learners often tuning out and not taking advantage of CAI’s abilities, even if they would have been useful for the learner’s goals. Initial engagement with CAI was slow, as students needed time to understand, explore, and feel comfortable with the new technology integrated into their learning environment and its interface.

5.7.3.2 Workflow adjustments

Additionally, we uncovered that prior EarSketch use seemed to make CAI adoption more difficult. We hypothesize that this is possibly due to the established mental models they had built with the EarSketch environment. Introducing CAI altered the students’ interaction dynamics with EarSketch, leading to a decline in engagement with other essential features related to code and music creation while they engaged with CAI.

5.7.3.3 Comparisons to established features

Students’ preferences leaned towards the more familiar and polished EarSketch over the experimental CAI, even though they expressed interest and appreciation for the latter’s capabilities.

5.7.4 Summer 2023: Summer Camp & Interviews

After the spring studies, we theorized that students’ previous experience with EarSketch diminished CAI’s effect beyond the opportunity cost of using CAI. In Summer 2023, we presented CAI to students without prior EarSketch experience and conducted interviews to gain more insight into how students adjust to CAI when learning EarSketch.

We studied CAI in a summer camp for high school students engaged in mathematics, science, and computing. The 13 summer camp students were split into two groups to complete a coding assignment: one group with assistance from CAI, while the remaining students completed an alternate task using the regular EarSketch interface. In addition to data logs and survey responses, students completed full interviews with researchers after the activity, discussing their experiences and demonstrating their created songs.

5.7.5 Findings

We performed a qualitative analysis of student interviews, focusing on questions about students’ experience working with CAI. Using inductive thematic analysis [16] and descriptive coding [148], followed by affinity mapping, we identified themes indicating that students experienced a mismatch in expectations of CAI vs. CAI’s actual behavior. This mismatch was largely due to previous experiences with recently released, accessible, flexible AI tools.

5.7.5.1 Varying expectations

Students had varying expectations for the system’s behavior. Many students expected CAI to help with “musical ideas,” “how to remix,” “setting mood” or “adding more beats and transitions.” Others expected primarily “coding help,” “more coding, less music” or for it to “write code or provide base code.” Among these students, one appreciated that CAI doesn’t code for them.

5.7.5.2 Experience with other AI applications

The advent of popular LLM applications informed subject expectations for CAI’s behavior. Of the 7 students who suggested open text as a potential improvement for the CAI system, 6 reported prior LLM chat tool use. Some students compared the purpose of CAI to that of

ChatGPT. Due to its targeted and specialized suggestions, one called it “more creative than ChatGPT.”

5.7.5.3 Understanding of CAI

Most students reported that the presence of CAI was helpful to them. Of the 13 students, 9 reported that CAI made it easier to find samples, with one stating that CAI helped “make the song smoother.” However, 5 students reported some “random” or “bad” samples. Due to a lack of explanation of the recommendation system’s generation of “contrasting” samples [163], these recommendations reduced subject trust in the system.

5.7.5.4 Time-saving strategies

Even students with less positive perceptions of CAI’s suggestions referred to it as a “safety net” that was useful when stuck and could add new ideas that the student would not have thought of. Many (7) students developed a similar strategy: they started by finding base sounds, genres, and instruments in the Sound Browser to save time by limiting CAI’s range of suggestions.

5.8 Key Takeaways for My Dissertation Research

The development and evaluation of CAI provided invaluable experience in studying human-AI creative collaboration. These findings highlight the importance of balanced interaction design, proactive support mechanisms, and careful integration with existing platforms. Additionally, the development and evaluation of CAI provided invaluable experience in studying human-AI creative collaboration. Through CAI’s development, I gained expertise in: (1) Designing and implementing controlled studies of human-AI interaction in creative tasks, including methods for capturing detailed interaction logs and analyzing creative processes (2) Understanding the importance of clear onboarding and expectation setting when introducing AI tools to users, particularly relevant given varying user familiarity with AI capabilities (3) Identifying and analyzing relevant features of creative artifacts produced through human-AI collaboration (4) Managing the practical challenges of conducting studies with AI tools, including technical infrastructure for data collection and analysis (5) Recognizing common patterns in how users approach AI assistance in creative tasks, such as their tendency to seek specific types of help

at different stages of the creative process. While my dissertation work uses an LLM rather than a menu-based system, these fundamental insights about studying human-AI creative collaboration remain highly relevant. The progression from studying structured interactions in CAI to examining more open-ended dialogue with LLMs represents a natural evolution in understanding how AI can effectively augment human creativity.

CHAPTER 6

INVESTIGATING HUMAN-AI CO-CREATIVITY IN NARRATIVE WRITING

6.1 Motivation and Research Questions

The overarching question of my dissertation is *how can AI support human creativity?* This question has guided my PhD journey. During the first three years of my PhD, I contributed to a team focused on developing a co-creative AI (CAI) partner for EarSketch, where students learn to code while creating music. My early research investigated human-human collaboration in creative tasks to understand the co-creative process (Chapter 3) and its relationship with partner satisfaction (Chapter 4). These investigations informed CAI's development, culminating in the analysis of our prototype's design and testing phases (Chapter 5). During this period, significant advancements in AI's creative capabilities have fundamentally shifted our understanding of human-AI creativity. These developments suggest a critical need to understand how these new AI capabilities influence the human creative process before we can effectively design systems to support it.

This dissertation focuses specifically on investigating the influence of conversational AI on narrative writing, both in terms of process and outcomes. The shift from my previous work to explore the impact of conversational AI on narrative writing is grounded in several key considerations. First, as revealed in the systematic literature review (Chapter 2), research on AI-augmented creativity shows an increasing emphasis on narrative tasks. By focusing on written narrative, this work builds upon and contributes to this expanding research direction. Second, while narrative writing demands its own sophisticated set of creative and compositional skills, these skills are more commonly developed through general education and daily communication than the specialized combination of coding proficiency and musical knowledge required in my previous studies. This broader prevalence of narrative writing skills increases the pool of potential participants, which aids recruitment efforts. Finally, research has shown that narratives are written with less effort, greater fluency, and greater coherence than other forms of writing [89], that creative writing processes exhibit faster writing, more stability, longer texts, and fewer revisions [178], and that open-ended, subjective writing activities are the most enjoyable [78]. These

attributes, combined with narrative writing's openness to imagination and subjective nature, provide an ideal environment for investigating how conversational AI can support and enhance human creativity.

To explore how conversational AI influences both the narrative writing process and final narrative artifacts, this study investigates two primary research questions:

RQ1 How does conversational AI assistance impact the users' final narrative artifacts in terms of text length, linguistic complexity, and cohesion compared to narratives written individually?

RQ2 How does conversational AI assistance influence users' creative decision-making and narrative development?

The rest of the chapter is organized as follows: Section 6.2 outlines the study methodology, including participant recruitment, procedure, and data collection methods. Section 6.3 describes the development of the web-based platform used to conduct this study. Sections 6.4 and 6.5 details the analysis and results of this analysis, and Section 6.6 discusses the implications of these findings for human-AI co-creativity.

6.2 Study Overview

This study¹ employs a within-subjects, randomized design to investigate how conversational AI influences narrative writing processes and outcomes. Participants completed two creative writing sessions—one with AI assistance and one individually—in a controlled laboratory environment. The study utilized a custom-built web platform to facilitate the writing tasks, track participant interactions, and collect data for subsequent analysis. Each session followed identical time constraints and task structures to enable direct comparison between conditions.

6.2.1 Participants and Recruitment

Initial pilot testing revealed systematic validity concerns regarding participant compliance with study protocols through online platforms, particularly in following instructions about external AI tool usage. Consequently, the recruitment strategy pivoted from its initial intent to in-person participation from university students.

¹ This study received exempt status approval from the University of Florida Institutional Review Board (Protocol #ET00045339).

While initial recruitment efforts focused on students in writing courses, I expanded my recruitment to include Computer and Information Science and Engineering (CISE) students to improve participation levels. In total, 25 students participated in this study. However, one participant was excluded from analysis after log data review revealed clear evidence of external AI usage during the individual condition (specifically, copying the prompt, leaving the study page, then returning to paste extensive pre-generated text). Additionally, one participant did not interact at all with the agent during the study and wrote both stories on their own. I removed both of these cases from my analysis, resulting in a final sample of 23 participants. I conducted the study sessions in a controlled laboratory environment on campus, which allowed for consistent administration of the research protocol and direct monitoring of participant engagement.

To maintain anonymity and encourage natural creative expression, I implemented a randomized assignment procedure with no collection of demographic data. I generated 120 generic usernames (in u### format) and passwords in a MySQL database, each randomly pre-assigned to either the AI-assisted or individual condition. During the study, participants selected a slip of paper containing login credentials after completing the oral consent process, with no indication of which condition each username was associated with. This procedure ensured that neither the participants nor I could identify which condition was assigned to specific individuals, and no connection existed between participants' identities and their study data beyond names recorded solely for course credit purposes. This resulted in 13 participants completing the AI-assisted condition first and 11 participants completing the individual condition first.

I managed participant recruitment and credit administration through two parallel processes: CISE students used the department's SONA system for both study sign-ups and participation reporting, while writing course students signed up through SONA but participation reporting required direct coordination between myself and their instructors.

The sample size ($n=23$ after exclusion) is smaller than the originally planned 100 participants when recruitment was intended through Prolific. This limitation, along with the

restricted demographic of university students, constrains the generalizability of findings that will be addressed in the discussion section.

6.2.2 Procedure

This study employs a within-subjects design with counterbalancing to compare how writing behaviors and outcomes differ with and without conversational AI assistance. During the study, participants engage in two writing sessions, one individually and one with AI assistance, each consisting of three main components:

1. A 4-minute character brainstorming activity (Appendix D)
2. A 16-minute story writing task (Appendix E)
3. A brief post-activity questionnaire (Appendix F)

In the AI-assisted condition, participants have access to conversational AI support within the study window during both the brainstorming and writing phases.

Following the completion of both writing sessions, participants complete several short questionnaires which ask participants to compare their experiences across both conditions (Appendix F), report level of agreement with statements assessing creative self-efficacy and creative personal identity (Appendix G) and statements assessing writing attitudes and writing self-efficacy (Appendix H), and provide information on their language background (Appendix I). Additionally, participants complete a modified large language model experience questionnaire to assess their familiarity with and usage of large language models (LLMs). The questionnaire examines experience with five LLMs: ChatGPT, Claude, Gemini, Grok, and other AI language models.² For each model, participants indicate their usage level on a seven-point scale ranging from “I do not know this model” to “Daily use.”

The study is designed to be completed within an hour to keep participants engaged and avoid fatigue. Progress tracking is implemented through a sidebar interface that displays

2 The original questionnaire [185] focused solely on ChatGPT. I expanded it to include other prominent LLMs available at the time of the study. However, DeepSeek is not included in this list as its public release coincided with the beginning of study sessions, making questionnaire modifications impractical at that stage

completed activities and upcoming tasks. A countdown timer is visible during timed activities to help participants manage their time effectively.

6.3 System Design and Development

To facilitate this study, I designed and implemented a custom web-based platform optimized for both individual and AI-assisted narrative writing tasks. The platform supported the complete research workflow, from participant authentication to data collection, while providing a consistent user experience across both experimental conditions.

6.3.1 Interface Design Process

My initial design process focused on the core study interfaces: the brainstorming and writing activity pages. Informed by familiar writing interfaces, such as Google Docs, I began with low-fidelity sketches in Goodnotes on my iPad, creating basic rectangular layouts divided into labeled sections for the key interface components, illustrated in Figure 6-1. For the Individual condition, these sketches outlined two main sections: a progress-tracking sidebar and writing area. The AI-assisted condition sketches included a third section for the chat interface. In both conditions, the writing area contained space for task instructions and writing input.

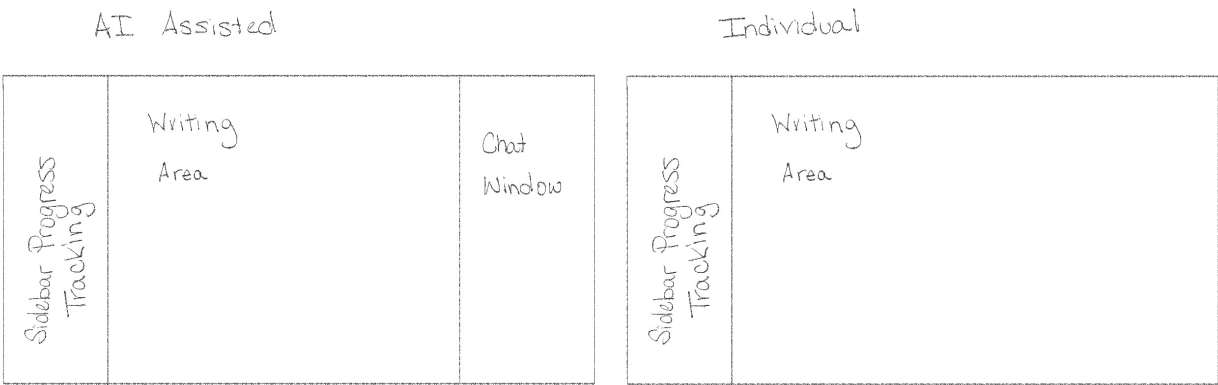


Figure 6-1. Initial low-fidelity sketches of the writing interface design showing the layout differences between AI-assisted (left) and Individual (right) conditions, with labeled sections for the sidebar progress tracking, writing area, and chat window.

After receiving feedback on the planned layout, I created a medium-fidelity Figma prototype based on the initial sketches that focused on general styling and component details. The prototype featured a dark theme with a white text editor, drawing from common commercial AI systems,

chat applications, and writing software. The progress tracking sidebar component displays activity timers during timed tasks and question counts for questionnaires, with a collapsible feature to minimize potential distraction.

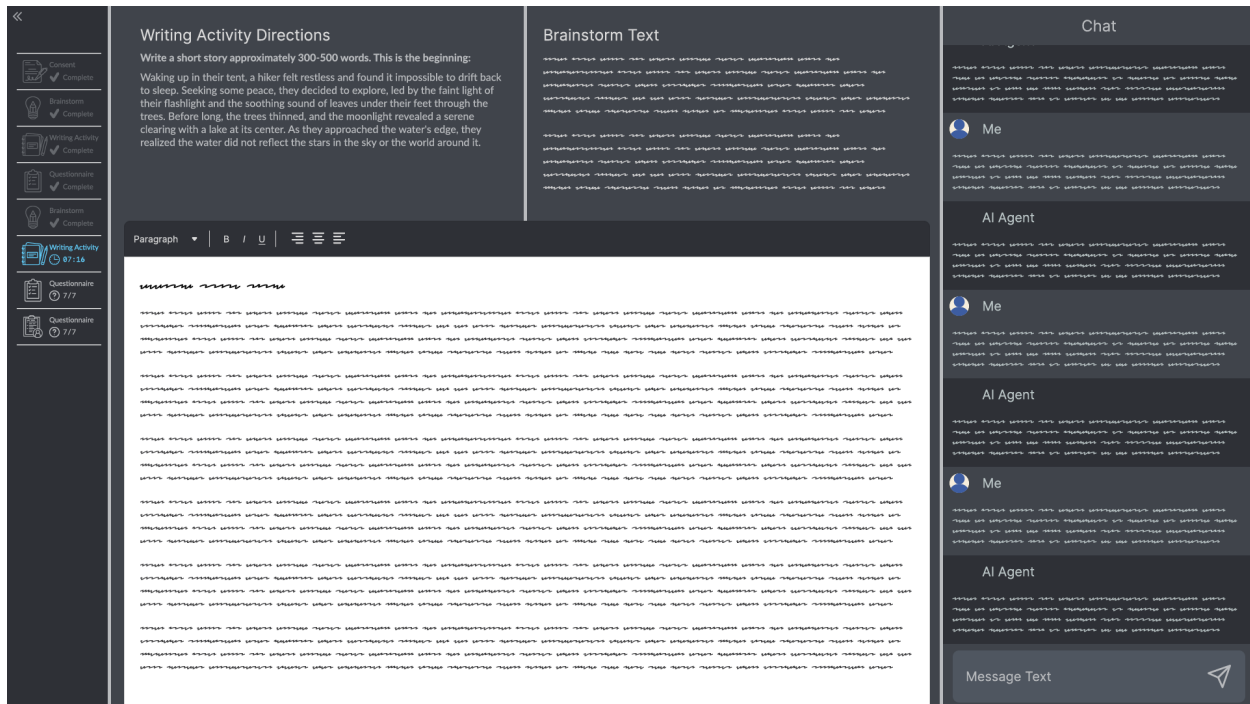


Figure 6-2. A medium-fidelity Figma prototype of the writing activity interface for the AI-assisted condition. The design features a progress-tracking sidebar (left) that displays task timers and question counts, a main writing panel (center) with activity instructions and participants' brainstorm text at the top and text editor below, and a chat window for AI interaction (right).

After receiving feedback on the Figma design, I began the web implementation of my study website. During this phase, I conducted user testing with friends, family, and co-workers, leading to four primary changes: First, I added an introduction page that informs participants of the study structure, technical requirements, and important procedural information. Second, I implemented a word count display below the text editor. Third, I added attention-notes to the brainstorm and writing activity pages that explicitly state whether participants should work independently or may use the AI assistance, emphasizing the prohibition of external resources in both conditions. Fourth, I revised the progress tracker from tab-like icons to a vertical sequence of round icons

connected by lines, providing participants with a clearer sense of progression through the study activities.

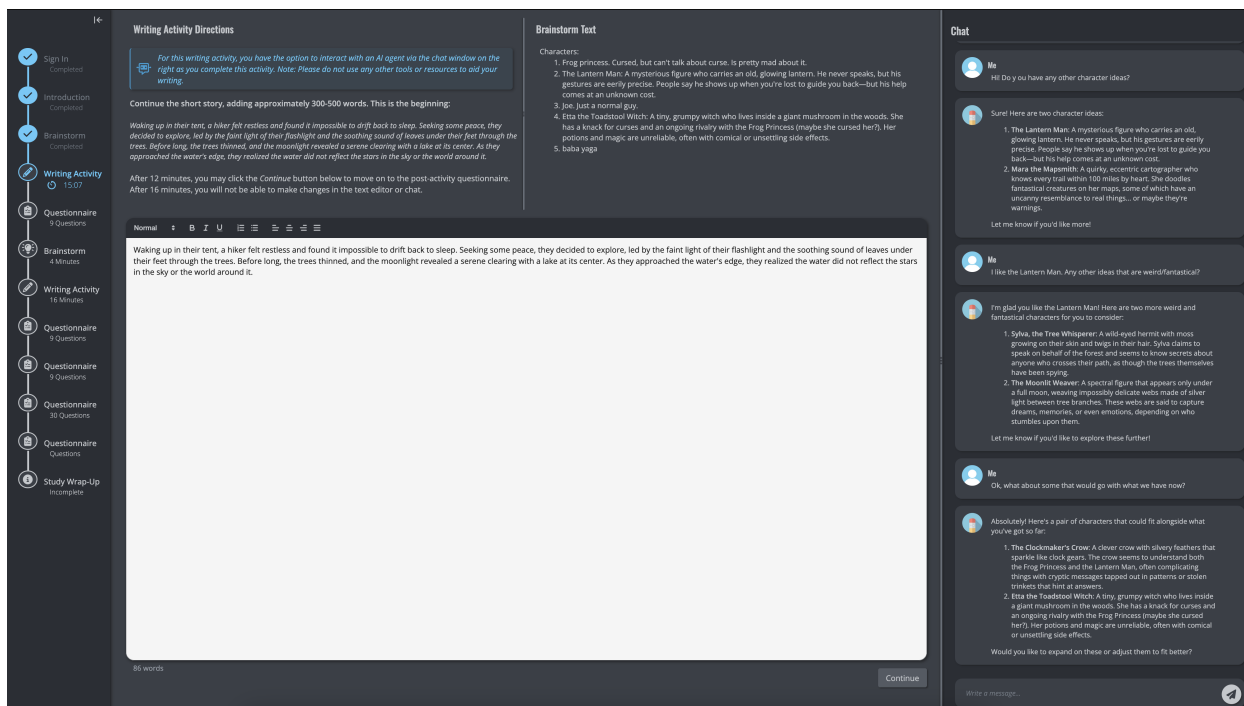


Figure 6-3. Final implemented web interface for the writing activity showing the vertical progress tracker (left), writing activity directions with prominent attention note (center-top), word count display (center-bottom) and chat interaction window (right).

6.3.2 Technical Architecture

The platform was built using a combination of front-end and back-end technologies:

- **Front-end:** HTML5, CSS3, and JavaScript provided the core structure, styling, and interactive elements of the interface. The Quill.js rich text editor library was implemented for both the writing area and chat interface, offering consistent text formatting capabilities while simplifying data capture.
- **Back-end:** PHP handled server-side logic, including authentication, data processing, and communication with external APIs. A MySQL database managed user credentials, interaction logs, questionnaire responses, and writing outputs.
- **AI Integration:** The AI agent was integrated through API interactions and response processing managed through PHP, with specific model details described in Section 6.3.5.

The platform used responsive design principles to ensure consistent experience across screen sizes, while optimizing for desktop use in the laboratory environment. Browser session

storage facilitated efficient data logging by minimizing server requests, with storage automatically cleared upon session completion.

6.3.3 User Interface Components

The interface consisted of three primary components designed to support the research objectives:

- **Writing Area:** A central text editor with formatting controls allowed participants to create and edit their narratives. The Quill.js editor provided a familiar word processor experience with essential formatting options. A word count display below the editor helped participants monitor their progress.
- **Chat Interface:** Available only in the AI-assisted condition, this component enabled participants to interact with the AI agent through a familiar messaging interface. All prompts and responses were timestamped and logged for analysis.
- **Progress Sidebar:** A persistent sidebar displayed the participant's progress through study activities, remaining time for timed tasks, and navigation controls. This element remained consistent across both conditions to maintain experimental control.

6.3.4 Logging and Data Capture

I captured data from multiple sources:

- **Questionnaire Responses:** Responses to all questionnaires (discussed in Section [6.2.2](#)) were captured and recorded in a MySQL database.
- **Event Logs:** The system captured interaction event records timestamped with millisecond precision. Logged events included keystroke data, copy/paste use and content, text editor content, final written artifacts, navigation events for the different areas (window/document, sidebar, text editor, and chat), button clicks, text selection, errors generated, and all messages submitted by participants and the responses received from the agent.
- **Conversation Logs:** From the event logs, I extracted the sent and received messages to create the conversational transcripts between participants and the agent.

Browser session storage was utilized to temporarily store interaction data during active writing sessions, minimizing the risk of data loss due to connection interruptions. This data was systematically transmitted to the server database at regular intervals and upon activity completion.

6.3.5 Agent Implementation

The AI-assisted condition features a chat-based agent using OpenAI's gpt-4o-2024-11-20 model. The agent is available to interact with through a chat window on the right side of the interface (Figure 6-3) during both the AI-assisted condition's brainstorming and writing activities.

To establish rapport and clarify the agent's role, I designed an initial introduction message that appeared at the beginning of each AI-assisted condition brainstorm activity: *"Hi! I'm an AI that will be joining you in this portion of the writing study. Feel free to bounce ideas off me or ask for suggestions. I am able to see what you've written and can help with brainstorming, character development, plot ideas, and more. How do you want to begin?"*

To support natural interaction, the system automatically sends the current text editor contents with every message exchange, enabling the agent to reference and respond to participants' evolving narrative without requiring manual context sharing. This design decision emerged from pilot testing, which revealed differences in participant expectations: those with limited LLM experience often assumed the agent would automatically "see" their writing because both elements appeared within the same interface, while participants familiar with commercial LLMs expected to manually provide context. Pilot testing demonstrated noticeable experience discrepancies when context wasn't automatically provided, yet no negative impact when participants redundantly shared context manually. Based on these observations, I implemented automatic context sharing that operated invisibly to participants—the shared context was filtered from displayed messages, maintaining a clean interface while ensuring consistent agent performance across all participants.

The agent maintained conversation continuity using the conversation history that carried over between activities. When participants transitioned from brainstorming to writing, the agent received: (1) the complete brainstorming text editor content, (2) new guidelines for the writing activity, and (3) the writing prompt and initial story beginning. This design ensured the agent remained aware of study progress and maintained relevant context throughout the creative process.

During pilot testing, I observed that the agent was prone to over-implementing the study's activity instructions, sometimes providing excessive content regardless of the participant's

specific request. To address this, I provided the agent with specific instructions for both the brainstorming and writing activities to ensure consistent, controlled behavior across participant interactions. These instructions were designed to balance supportive collaboration while preventing the agent from dominating the creative process.

For both activities, the agent was instructed to wait for explicit requests before offering suggestions, tailor responses to the specific question or request made by the user, and follow all study directions. The agent was also provided with context about the study (a collaborative creative writing study on how humans write with AI assistance), its role (to provide support), and a copy of the current activity's complete instructions. Additionally, it was informed that it will receive text editor content updates for context.

During the brainstorming activity, the agent was instructed to limit the character suggestions to a maximum of two at a time. During the writing activity, the agent was instructed to only generate a single paragraph at a time if the user requests content. For non-narrative requests (character development, world-building, plot ideas), provide focused responses without extending the story. For both activities, the agent was informed that explicit user requests contrary to these guidelines should take precedence.

These guidelines were designed to be flexible rather than restrictive. Participants who desired more generated content could simply request it, while those who preferred brief responses were not overwhelmed with lengthy text. This design decision prevented situations where participants asking a simple question would receive extensive narrative content due to the agent's interpretation of the activity instructions. The balanced approach created a supportive writing partner that would not dominate the creative process—aligning directly with this dissertation's focus on investigating how AI can support human creativity.

The agent interface was designed with a focus on usability and unobtrusive integration with the writing environment. A few key elements include: A dedicated chat panel positioned on the right side of the writing interface; Visual differentiation between user and AI messages, with profile images for easy identification; A typing indicator showing when the agent was preparing a

response; Markdown rendering for agent responses to improve readability; Resizable panels allowing participants to adjust the chat window based on preference; Automatic scrolling to keep the most recent messages visible.

6.4 RQ1 Analysis and Results

This section details the methods used to examine the impact of conversational AI assistance on narrative writing, along with their corresponding results. The methods and results are organized according to the specific hypotheses tested, covering both quantitative measures for analyzing narrative artifacts (H1-H5, addressing RQ1) and mixed-methods approaches for understanding the creative process (addressing RQ2).

For most analyses related to RQ1 (H1-H4), I employ a within-subjects design, comparing each participant's performance across the AI-assisted and individual writing conditions. This approach controls for individual differences in writing ability, creative tendencies, and other participant characteristics that might influence the outcomes.

The semantic homogeneity analysis (H5, also addressing RQ1) follows a different approach, as it examines similarities between narratives within each condition rather than comparing each participant's work across conditions. For this analysis, I use independent statistical tests to compare similarity distributions between conditions.

6.4.1 H1: Text Length

This hypothesis examines whether AI assistance affects the basic production metrics of narrative writing. By comparing word and character counts between conditions, this analysis investigates whether conversational AI support influences the quantity of text participants produce.

6.4.1.1 Analysis

I hypothesize that narratives produced with conversational AI assistance would be longer than those produced individually, based on the AI's potential to assist with idea generation and text production. To test this hypothesis, I compared both word count and character count across conditions. I initially planned to use paired t-tests, but after examining the distribution, shown in Figure 6-4, and conducting a Shapiro-Wilk test, which indicated a significant deviation from

normality ($p < .001$), I employed the non-parametric Wilcoxon signed-rank test instead. For each analysis, I calculated standardized effect sizes using the formula $r = Z/\sqrt{N}$ [147], where Z is the standardized test statistic and N is the total number of observations. This standardized effect size measure allows for interpreting the magnitude of differences according to Cohen's guidelines [29], with values of 0.1, 0.3, and 0.5 representing small, medium, and large effects, respectively.

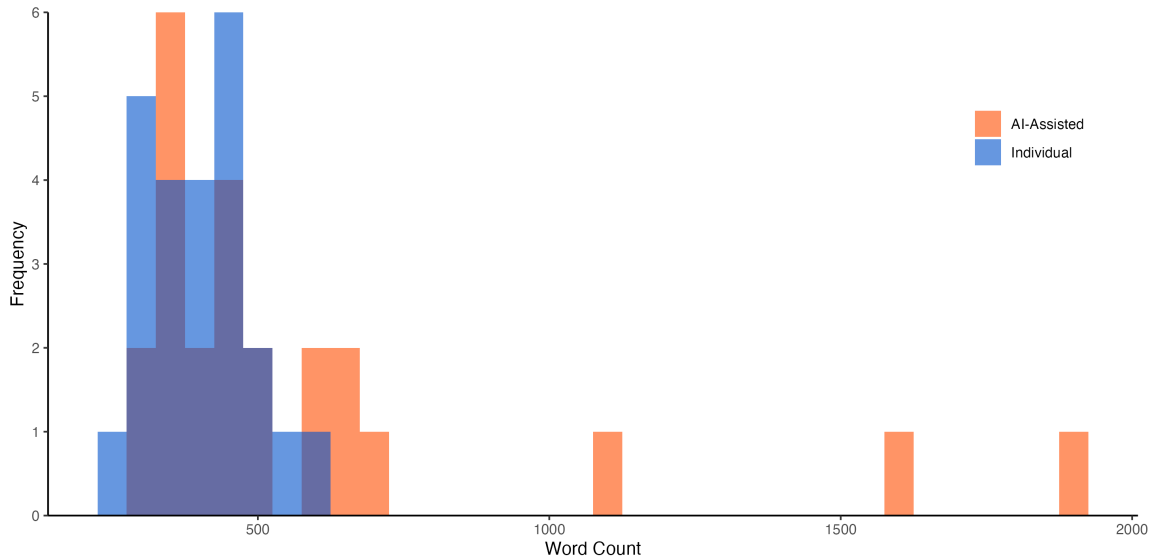


Figure 6-4. Distribution of word counts across experimental conditions

6.4.1.2 Results

Analysis of word counts revealed that texts produced with AI assistance contained significantly more words ($M = 584.25$, $SD = 406.26$) than those written individually ($M = 399.29$, $SD = 92.21$). The exact Wilcoxon signed-rank test confirmed this difference ($V = 228$, $p = .012$), with a median increase of 69 words when using AI assistance. The effect size ($r = 0.452$) indicates a medium effect according to Cohen's guidelines [29].

Character count analysis similarly demonstrated that AI-assisted texts contained significantly more characters ($M = 3393.08$, $SD = 2505.67$) than individually written texts ($M = 2191.17$, $SD = 522.92$). The exact Wilcoxon signed-rank test confirmed this difference ($V = 240$, $p = .005$), with a median increase of 371 characters when using AI assistance. The effect size ($r = 0.522$) indicates a large effect according to Cohen's guidelines [29].

The distribution of difference scores departed substantially from normality (Shapiro-Wilk test: word count $W = 0.694$, $p < .001$; character count $W = 0.679$, $p < .001$). This non-normal distribution was largely driven by three participants (12.5% of the sample) who produced exceptionally longer texts in the AI-assisted condition. These participants showed a word increase from Individual to the AI-assisted condition 11-20 times larger than the median increase of 69 words, with similar trends for character count.

These findings support the hypothesis that participants produce longer texts when writing with AI assistance compared to writing individually. The substantial variance in the AI-assisted condition (word count variance ratio = 19.41, character count variance ratio = 22.96) suggests that AI's impact on text production varies considerably across individuals, with some participants leveraging AI to produce substantially longer narratives while others showed more modest increases.

6.4.2 H2: Linguistic Features

We investigate whether AI assistance affects the linguistic characteristics of narratives. Through analysis of lexical richness and syntactic complexity measures, this section examines how AI collaboration may alter the linguistic sophistication and structural properties of narrative writing.

6.4.2.1 Analysis

I hypothesize that narratives produced with conversational AI assistance exhibit distinct linguistic characteristics compared to those produced individually, particularly regarding lexical richness and syntactic complexity. Given the recent emergence of conversational AI models like GPT-4, I adopted an exploratory approach using established measures from second-language (L2) research and writing studies.

To investigate the impact of AI assistance on linguistic complexity and sophistication in narrative writing, I analyzed seven linguistic features across two main categories: lexical richness and syntactic complexity. All measures were min-max scaled within each feature type to range from 0 to 1, enabling direct comparisons across different linguistic dimensions despite their

varying original scales. All linguistic features were calculated using NeoSCA [174]. The abbreviations used for the features matches those in the NeoSCA software.

Lexical richness measures: I examined three dimensions of lexical richness: Mean Segmental Type-Token Ratio, Lexical Density, and Lexical Sophistication.

Mean Segmental Type-Token Ratio (MSTTR) quantifies lexical diversity by calculating the ratio of unique words to total words across text segments. I selected this measure for its reduced sensitivity to text length [109, 12], with higher values indicating greater vocabulary diversity.

Lexical Density (LD) represents the proportion of content words (nouns, verbs, adjectives, and adverbs) to total words, reflecting the informational density of the text.

Lexical Sophistication (LS1) measures the proportion of sophisticated vocabulary items in the text by comparing word usage against the American National Corpus [140], with higher values suggesting more advanced word choices.

Syntactic complexity measures: I examined four dimensions of syntactic complexity: Mean Clause Length, Clauses per T-unit, Coordinate Phrases per Clause, and Complex Nominals per Clause.

Mean Clause Length (MLC) calculates the average number of words per clause, with longer clauses typically indicating more complex syntactic structures. A clause is a group of words containing a subject and a predicate (verb) that functions as a unit within a sentence, either independently as a simple sentence or dependently within a complex sentence. This measure captures the length dimension of language units.

Clauses per T-unit (C/T) measures the ratio of clauses to T-units. A T-unit is a main clause plus any subordinate clauses or non-clausal structures attached to or embedded within it, representing the minimal terminable unit of language capable of standing alone as a grammatical sentence. C/T represents sentence complexity through subordination.

Coordinate Phrases per Clause (CP/C) measures the use of coordinate phrases per clause. A coordinate phrase is a grammatical construction that joins two or more elements of equal

syntactic rank or importance using coordinating conjunctions (such as “and,” “or,” “but”), creating a compound structure that functions as a single unit within a sentence. CP/C represents complexity through coordination at the phrasal level.

Complex Nominals per Clause (CN/C) counts the frequency of complex noun phrases per clause. A complex nominal is a noun phrase consisting of a head noun modified by adjectives, possessives, prepositional phrases, relative clauses, participles, or appositives; it also includes nominal clauses, gerunds, and infinitives in subject position. CN/C represents sophistication in nominal structures and phrasal complexity.

To analyze the differences between AI-assisted and individual writing conditions, I conducted a repeated measures multivariate analysis of variance (MANOVA) in R using the *afex*³ library, followed by post-hoc analyses for each measure using the *emmeans*⁴ library. This approach allowed me to examine both the overall impact of AI assistance on linguistic features and its specific effects on individual measures while controlling for the relationships between different linguistic dimensions. Because linguistic features can vary based on language background, I included participants’ first language status (English as a first language vs English as an additional language) as a control variable in my analysis.

6.4.2.2 Results

I first ran a multivariate analysis of variance (MANOVA) with condition (AI-assisted vs. individual) as a within-subjects factor and language background (English as a first language vs English as an additional language) as a between-subjects factor. The MANOVA revealed a significant main effect of condition ($p = .025$) and a significant condition by measure interaction ($p < .001$, Greenhouse-Geisser corrected). Neither the main effect of language background nor any of its interactions with other factors were significant (all $p > .17$). Therefore, while language background was included in the model, it is not discussed further in the analysis of condition effects. The complete MANOVA results are presented in Table 6-1.

3 <https://cran.r-project.org/package=afex>

4 <https://cran.r-project.org/package=emmeans>

Table 6-1. Results of MANOVA with Condition (AI-Assisted vs. Individual) and Language Background as Factors

Effect	df	F	η_G^2	p-value
is_english.l1	1, 22	0.08	<.001	.779
condition	1, 22	5.80	.031	.025*
is_english.l1:condition	1, 22	1.96	.011	.175
measure	3.74, 82.19 ^a	24.79	.325	<.001***
is_english.l1:measure	3.74, 82.19 ^a	0.98	.019	.421
condition:measure	2.59, 56.96 ^a	7.58	.086	<.001***
is_english.l1:condition:measure	2.59, 56.96 ^a	0.65	.008	.563

Note: ^aGreenhouse-Geisser correction applied due to sphericity violation.

* $p < .05$, ** $p < .01$, *** $p < .001$

Following the significant MANOVA results, I conducted post-hoc tests using estimated marginal means to determine which specific linguistic features were affected by AI assistance. Two lexical measures showed significant differences between conditions: 1) Mean Segmental Type-Token Ratio (MSTTR) was significantly higher ($p = .001$) for texts written in the AI-assisted condition ($M = 0.706$) compared to individually written texts ($M = 0.478$). This indicates that AI-assisted texts exhibited greater lexical diversity within segments, suggesting less word repetition. 2) Lexical Sophistication (LS1) was significantly higher ($p < .001$) in AI-assisted texts ($M = 0.623$) than in individually written texts ($M = 0.297$). This finding indicates that AI-assisted texts contained a higher proportion of sophisticated or less common vocabulary. Both of these effects remain statistically significant even with a conservative Bonferroni correction for multiple comparisons (adjusted $\alpha = .007$).

No significant differences were found in Lexical Density ($p = .387$). Similarly, none of the syntactic measures showed significant differences between conditions: Mean Clause Length ($p = .844$), Coordinate Phrases per Clause ($p = .780$), Complex Nominals per Clause ($p = .241$), and Clauses per T-unit ($p = .966$), as detailed in Table 6-2.

These results partially support my second hypothesis: AI assistance did significantly affect some aspects of linguistic complexity, specifically leading to increased lexical diversity and increased lexical sophistication. The magnitude of these differences was substantial: lexical sophistication showed the largest effect, with AI assistance increasing LS1 by 32.6% of the

Table 6-2. Univariate Comparisons and Descriptive Statistics of Linguistic Features by Condition

Measure	AI-Assisted		Individual		Difference	p-value
	Mean	SD	Mean	SD		
Mean Segmental Type-Token Ratio (MSTTR)	0.706	0.190	0.478	0.216	0.228	.001**
Lexical Sophistication (LS1)	0.623	0.244	0.297	0.188	0.326	<.001***
Lexical Density (LD)	0.607	0.200	0.575	0.222	0.032	.387
Mean Clause Length (MLC)	0.444	0.213	0.480	0.277	-0.036	.844
Coordinate Phrases per Clause (CP/C)	0.432	0.260	0.434	0.265	-0.003	.780
Complex Nominals per Clause (CN/C)	0.309	0.130	0.385	0.229	-0.076	.241
Clauses per T-unit (C/T)	0.130	0.192	0.135	0.120	-0.004	.966

Note: Measures are sorted by descending mean values for the AI-Assisted condition. All measures are min-max scaled to a 0-1 range.

* $p < .05$, ** $p < .01$, *** $p < .001$

possible range, while MSTTR increased by 22.8% of its possible range. However, the results do not show evidence of the AI assistance impacting syntactic complexity or lexical density.

6.4.3 H3: Referential Cohesion

This hypothesis explores whether AI assistance affects how ideas are connected through reference across a narrative. By analyzing referential cohesion scores, this section investigates whether AI support changes how effectively entities and concepts are incorporated throughout texts.

6.4.3.1 Analysis

I hypothesize that narratives produced with AI assistance would show different referential cohesion patterns compared to those written individually. Referential cohesion is a measure of text readability. It refers to the extent to which ideas in a text are connected through overlapping words, concepts, or referents across sentences and paragraphs, which helps readers track entities and concepts throughout a narrative. To analyze referential cohesion in participants' narratives, I used Coh-Metrix to calculate standardized referential cohesion scores (PCREFz)[119].

Initially, I planned to conduct a paired t-test to compare referential cohesion scores between AI-assisted and individually written narratives. However, research has established that text length and cohesion measures are related, as increasing cohesion typically requires adding words to fill conceptual gaps between ideas [118]. Given the findings from H1 (Section 6.4.1) that

demonstrated significant differences in text length between conditions (with AI-assisted texts being substantially longer), I first needed to determine whether including text length was appropriate for this analysis. To address this, I compared two mixed-effects models using the *lme4*⁵ package: a basic model with only condition as a fixed effect and participant ID as a random effect and a model that added word count as a predictor. I used a likelihood ratio test to determine whether accounting for text length significantly improved model fit. The likelihood ratio test indicated that adding word count and its interaction with condition significantly improved model fit ($\chi^2 = 6.33$, $df = 2$, $p = .042$), confirming that text length should be accounted for when analyzing referential cohesion. Based on the model comparison result, I proceeded with the mixed-effects model that included condition, word count, and their interaction as fixed effects, with participant ID as a random effect.

6.4.3.2 Results

To examine how AI assistance affects referential cohesion, I compared referential cohesion scores of the final narrative artifacts. Initial descriptive statistics showed that texts produced with AI assistance had lower mean referential cohesion scores ($M = -0.629$, $SD = 0.739$) compared to individually written texts ($M = 0.008$, $SD = 1.08$).

Based on the model comparison result discussed above, I proceeded with the mixed-effects model that included condition, word count, and their interaction as fixed effects, with participant ID as a random effect. The fixed effects coefficients showed a significant positive effect for the individual condition ($\beta = 2.16$, $p = .009$), and significant negative interaction between individual condition and word count ($\beta = -0.004$, $p = .034$).

Based on the results of the mixed-effects model, I derived an ANOVA table from the same mixed-effects model using the *anova()* function from the *lmerTest*⁶ package. These two outputs from the same model offer complementary perspectives: the mixed-effects model coefficients show the estimated effects at specific reference levels while the ANOVA tests whether each factor

5 <https://cran.r-project.org/package=lme4>

6 <https://cran.r-project.org/package=lmerTest>

as a whole contributes significantly to explaining variance in referential cohesion while controlling for all other terms in the model.

The ANOVA results (Table 6-3) confirmed significant effects of condition ($F(1, 36.96) = 7.67, p = .009$), word count ($F(1, 43.25) = 6.76, p = .013$), and their interaction ($F(1, 39.07) = 4.86, p = .033$). While the main effects indicate that both condition and word count independently influence referential cohesion, the significant interaction qualifies these effects by demonstrating that the influence of condition depends on text length.

Table 6-3. Type III ANOVA Results for Referential Cohesion Mixed-Effects Model

Effect	Sum Sq	F	df	p-value
Condition	3.507	7.666	1, 36.96	0.009**
Word count	3.093	6.762	1, 43.25	0.013*
Condition × word count	2.224	4.862	1, 39.07	0.033*

Note: * $p < .05$, ** $p < .01$.

The model demonstrated good explanatory power, with fixed effects alone explaining 21.7% of variance (marginal R^2) and 50.2% when including random effects (conditional R^2). These results support the hypothesis that AI assistance alters referential cohesion in narrative artifacts.

6.4.4 H4: Deep Cohesion

This hypothesis examines whether AI assistance influences the logical and causal connections within narratives. Through analysis of deep cohesion scores, I investigate whether AI collaboration affects the extent to which relationships between events and concepts are explicitly signaled in text.

6.4.4.1 Analysis

I hypothesize that narratives produced with AI assistance would show different deep cohesion patterns compared to those written individually. Deep cohesion refers to the extent to which a text contains explicit causal and logical connectives that help readers form causal connections between ideas and events in a text. These connectives (such as “because,” “therefore,” and “consequently”) signal relationships between concepts and events, facilitating readers’ comprehension by making the logical structure of the narrative more transparent. To

analyze deep cohesion in participants' narratives, I used Coh-Metrix to calculate standardized deep cohesion scores (PCDCz)[119].

I initially planned to conduct a paired t-test to compare deep cohesion scores between AI-assisted and individually written narratives. However, given the findings from the referential cohesion analysis (Section 6.4.3) indicated a significant relationship between text length and referential cohesion, I first needed to determine whether text length was also a significant factor for deep cohesion. To address this, I compared two mixed-effects models: a basic model with only condition as a fixed effect and participant ID as a random effect and a model that added word count as a predictor. I used a likelihood ratio test to determine whether accounting for text length significantly improved model fit. The likelihood ratio test indicated that adding word count as a predictor did not significantly improve model fit ($\chi^2 = 1.06$, $df = 2$, $p = .59$), so I proceeded with the paired t-test as originally planned.

6.4.4.2 Results

To investigate the impact of AI assistance on deep cohesion in narrative writing, I compared standardized deep cohesion scores between AI-assisted and individually written texts. Initial descriptive statistics showed that texts produced with AI assistance had slightly lower mean deep cohesion scores ($M = -0.059$, $SD = 0.587$) compared to individually written texts ($M = -0.035$, $SD = 0.615$), though this difference was minimal.

Based on the model comparison result discussed above, I conducted a paired t-test of the deep cohesion scores between the texts written in each condition. The results revealed no significant difference ($t(22) = -0.120$, $p = .906$) in deep cohesion between AI-assisted texts and individually written texts, with a mean difference of -0.024 (95% CI [-0.433, 0.386]). These results do not support the hypothesis that AI assistance significantly affects deep cohesion patterns in narrative writing.

6.4.5 H5: Semantic Homogeneity

This hypothesis investigates whether AI assistance influences the semantic diversity across different writers' narratives. By measuring semantic similarity between texts, this analysis examines whether AI support leads to greater conceptual homogenization.

6.4.5.1 Analysis

I hypothesize that narratives created with AI assistance demonstrate greater semantic similarity to one another than those written individually. To test this hypothesis, I conducted a semantic homogeneity analysis comparing texts produced under the AI-Assisted and the Individual writing conditions.

For each writing activity, participants were given the beginning of a story and asked to complete it. While these prompts facilitated the writing task, their inclusion would introduce artificial similarities between texts. Therefore, I removed these prompts while preserving any participant modifications.

The prompt removal algorithm identifies prompt text by comparing the beginning of each story to the original prompt using cosine similarity, applies a threshold approach to accommodate cases where participants modified portions of the prompt, and determines the endpoint using a sliding window technique.

To quantify semantic homogeneity across narratives, I employ the Sentence-T5-xxl model[130], which captures semantic meaning through contextual embeddings. Using this model, I generated high-dimensional vector representations (embeddings) for each narrative. These embeddings capture the semantic content of the texts in a way that allows for meaningful comparison through vector operations. The model encodes not just lexical features but semantic and contextual information as well.

For similarity calculations, I compared each narrative to every other narrative within its condition (AI-Assisted or Individual) using cosine similarity to generate scores ranging from 0 (completely dissimilar) to 1 (identical). This process resulted in a similarity matrix for each condition displaying all pairwise comparisons.

Next, I extracted unique similarity scores from each matrix, using only the upper triangle to avoid redundant comparisons. I conducted a one-tailed *t*-test to determine whether the mean similarity score was significantly higher in the AI-Assisted condition compared to the Individual condition.

This approach, illustrated in Figure 6-5, enables me to evaluate whether AI assistance leads to greater semantic similarity across different writers' narratives, potentially indicating a homogenizing effect of AI tools on creative writing.

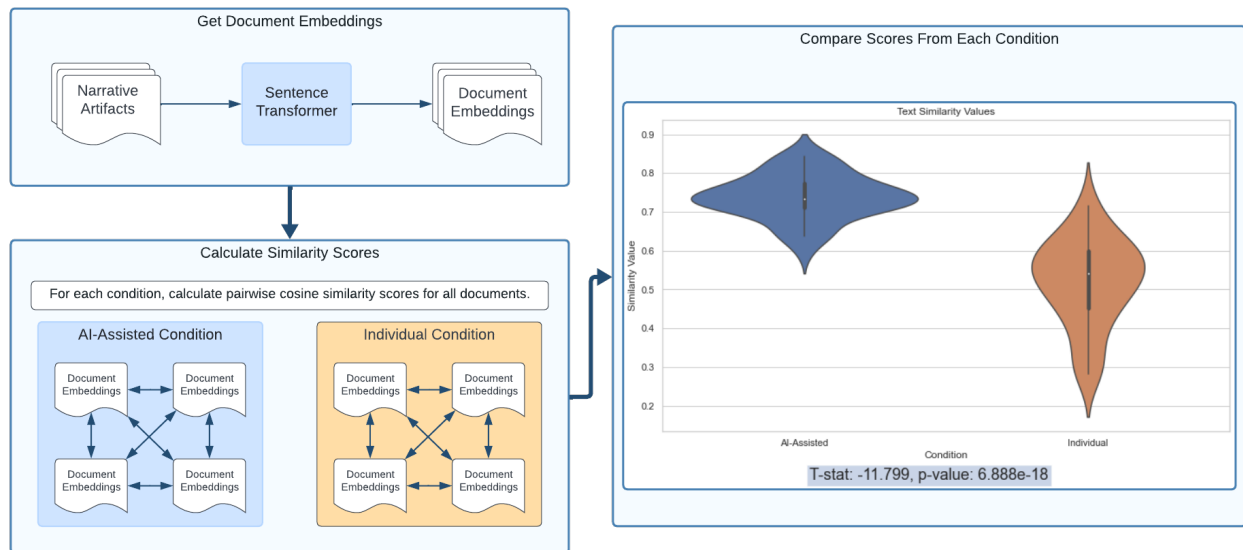


Figure 6-5. Methodological flowchart for assessing semantic homogeneity in narratives. Begins with generating the document embeddings, followed by calculating the within-condition pairwise cosine similarity scores, then ends with the one-tailed *t*-test (shown at the bottom of the violin plot).

6.4.5.2 Results

The semantic similarity analysis reveals significant differences in the homogeneity of narratives between the AI-Assisted and Individual writing conditions.

Before conducting inferential statistical tests, I verified that the data met the necessary assumptions. The Shapiro-Wilk test indicated that similarity scores were normally distributed in both the AI-Assisted ($W = 0.995$, $p = 0.606$) and Individual conditions ($W = 0.994$, $p = 0.425$). Additionally, Levene's test confirmed homogeneity of variances between the two conditions ($F = 1.36$, $p = 0.244$).

The descriptive statistics showed that narratives in the AI-Assisted condition had higher mean semantic similarity scores ($M = 0.879$, $SD = 0.029$) compared to narratives in the Individual condition ($M = 0.832$, $SD = 0.028$). This indicates greater semantic homogeneity among AI-assisted texts than individually written texts. Figure 6-6 shows the frequency of pairwise similarity scores, with density curves overlaid. AI-Assisted narratives demonstrate consistently higher similarity scores compared to Individual narratives, with minimal overlap between distributions, indicating a clear shift toward greater semantic homogeneity with AI assistance.

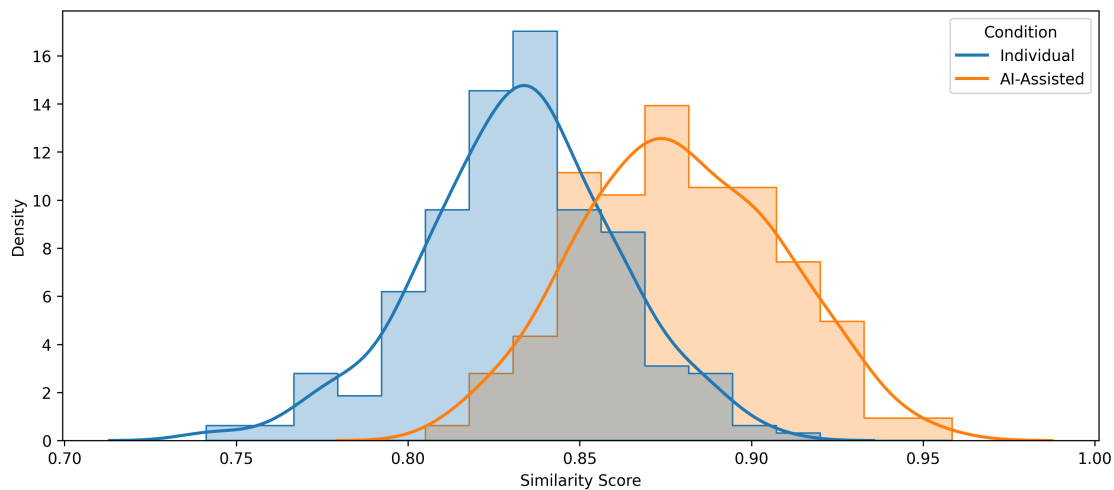


Figure 6-6. Distribution of semantic similarity scores for narratives in the Individual (blue) and AI-Assisted (orange) conditions. The histogram

An independent samples t-test confirmed that this difference was statistically significant ($t(504) = 18.38$, $p < .001$), with AI-assisted narratives exhibiting substantially higher semantic similarity to one another than individually written narratives. The effect size was very large (Cohen's $d = 1.64$), indicating a strong and meaningful difference in semantic homogeneity between conditions.

These findings support the hypothesis that AI assistance leads to greater semantic homogeneity across different writers' narratives.

6.5 RQ2 Analysis and Results

RQ2 employs a mixed-methods approach to understand how writers' interactions with conversational AI affects creative decision-making and narrative development. The analysis combines qualitative coding of participant utterances with quantitative insights from interaction logs and questionnaire responses.

The results presented in this section evolved from what I initially proposed for RQ2. Before presenting the results, I feel it is worthwhile to tell the “story” of how this direction emerged within my analysis.

My initial proposal for RQ2 focused on analyzing the narrative development process: comparing how linearly participants wrote their stories, examining revision patterns, tracking when they introduced characters and the consistency of characters from brainstorming to the final narrative, and identifying how settings and events were introduced between conditions. Additionally, I proposed analyzing how the interactions with the AI agent may have shaped the narrative elements and which components participants delegated to the AI agent.

The investigation began with the content attribution mapping (described in Section 6.5.1) to provide a foundation for distinguishing between participant-authored and AI-authored narrative content as well as ideational origins of the narrative content. Concurrently, I labeled the participants' messages for dialogue act analysis (described in Section 6.5.2) to provide a systematic framework for examining interaction patterns and how they shaped the artifact creation.

However, an unexpected finding emerged during my preliminary analysis: the polarization in how participants distributed creative agency between themselves and the AI agent. The extreme polarization appeared to be a fundamental aspect of human-AI co-creativity that required deeper investigation.

Consequently, the focus of the analysis shifted to exploring the relationship between dialogue patterns and creative agency distribution. Therefore, the results presenting in the following sections address my research question about how AI assistance influences creative decision-making, but through an emergent lens that focuses on creative agency distribution rather

than narrative construction processes. This approach provides unexpected but valuable insights into the mechanism through which humans delegate creative responsibility with AI collaborators.

6.5.1 RQ2: Mapping Participant-AI Content Attribution

This analysis explores how participants incorporate AI contributions into their final narrative artifacts. By mapping both the ideational and textual origins of the artifact content, this section reveals how creative agency and responsibility are distributed between human writers and the AI agent during collaborative writing.

6.5.1.1 Analysis

To understand how participants incorporate AI contributions into their narrative artifacts, I conduct a dual-dimension co-creative attribution analysis that traces both the ideational and textual origins in the final artifacts. Each word in the artifacts is coded according to two dimensions: textual authorship (participant-authored text versus AI-authored text with minimal modification) and ideational origin (participant-originated concepts versus AI-originated concepts).

For each participant, I calculate the percentage distribution across both dimensions: participant-authored versus AI-authored words, and participant-originated versus AI-originated ideas. Based on the relative proportions across these dimensions, participants are categorized into one of four co-creative patterns: **High Agency** (predominantly participant ideas and participant text); **High Delegation** (predominantly AI ideas and AI text); **Expressive Delegation** (predominantly participant ideas expressed through AI text); and **Conceptual Delegation** (predominantly AI ideas expressed through participant text).

6.5.1.2 Results

The attribution mapping reveals distinct interaction patterns in how participants engage in co-creative writing with AI, as shown in Figure 6-7. The majority of participants (12 of 23; 52.2%) cluster in the High Agency quadrant, maintaining primary authorial and ideational control. A substantial minority (9 of 23; 39.1%) fall into the High Delegation quadrant, where creative agency was largely transferred to the AI system for both ideation and textual expression.

The remaining participants (2 of 23; 8.7%) display Expressive Delegation patterns—contributing their own ideas while leveraging AI-generated text as an expressive medium.

Notably, no participant occupies the Conceptual Delegation quadrant (using predominantly AI-originated ideas while maintaining textual authorship), suggesting a hierarchy in creative delegation: participants appear willing to delegate expressive tasks but are reluctant to relinquish control over conceptual direction while assuming responsibility for writing. This suggests a potential asymmetry in how creative agency is negotiated in human-AI collaborative writing.

The distribution reveals that participants generally either maintain high creative control or engage in substantial delegation to the AI, with few adopting hybrid approaches. The High Agency quadrant displays greater variation in attribution patterns compared to the High Delegation quadrant. Nine participants' artifacts consist entirely of AI-authored text (excluding study prompt), with five of those also having 100% AI-derived ideational content. These findings demonstrate significant variation in the distribution of creative agency between human and AI, ranging from minimal AI influence to complete reliance on AI-generated material.

Figure 6-8 illustrates the sequential distribution of attribution throughout participants' final artifacts from the AI-assisted condition. Each horizontal bar represents a single document, with the width of colored segments proportional to their percentage of the total document length. The visualization reveals distinctive patterns in interaction provenance: Study Text (gray) appears at the beginning of documents. Participant-authored text (green) is prevalent across most artifacts, while AI contributions appear in different forms—either as direct AI-authored text (purple), participant-modified AI ideas (blue), or AI implementations of participant ideas (orange). Notably, because all participants received the same initial story prompt text, the relative width of these gray segments provides an indication of the overall document length where narrower gray segments indicate longer total documents.

6.5.2 RQ2: Dialogue Act Analysis

This analysis examines the patterns of interaction between participants and the AI agent during the writing process. By labeling and analyzing the dialogue acts, this section reveals how

participants formulate requests, integrate AI suggestions, and delegate creative decisions through conversation.

6.5.2.1 Analysis

To understand how participants engage with the AI assistant during narrative construction, I conduct a dialogue act analysis of the conversational interactions, with a particular focus on how participants formulate requests and integrated AI responses into their writing process.

I extract all conversation logs, resulting in 266 total messages, with 133 participant messages across 23 participants. I independently conduct an initial round of open coding on the conversation data, identifying and labeling interaction patterns with particular attention to process elements in how participants engaged with the AI assistant. Drawing on my initial codes and relevant literature in dialogue act theory, collaborative writing, and conversational UX design, I developed a coding taxonomy to classify participant utterances. A second researcher and I reviewed and applied the proposed taxonomy. Given the relatively small dataset, the second researcher and I collaboratively coded all 133 participant messages. When coding disagreements arose, we discussed the rationale behind our different interpretations, ultimately reaching consensus and refining the taxonomy definitions for clarity.

The final coding taxonomy organizes participant interactions through a hierarchical structure. At the top level, we categorized each message by its primary illocutionary act: Request, Clarification/Elaboration, Assessment, or Other (Table J-1). For messages coded as Requests, we applied additional subcodes (Table J-2) to capture their specific characteristics, discussed below.

Request Type: Distinguishes whether participants sought content generation (asking the agent to generate or modify text), information-seeking acts (requesting factual information or explanations), or evaluative queries (soliciting assessment of existing content).

Request Directness: Represents the dialogue act form, differentiating between direct dialogue acts (where the illocutionary force is explicitly conveyed through linguistic form) and indirect dialogue acts (where the illocutionary force requires pragmatic inference beyond the semantic meaning).

Request Constraint: Represents the degree of constraint participants provided in their directives, classifying them as bounded (specific parameters that constrain the AI's response space) or unbounded (minimal parameters allowing broader creative latitude).

Request Focus: Represents the subject matter focus of each request, such as character development, plot advancement, setting description, thematic exploration, stylistic concerns, structural elements, or informational aspects.

Request Iteration. Represents whether directives are new requests, repeated previous requests, or refined earlier requests with additional or modified specifications.

Content Operation: (For content generation requests specifically) Represents the relationship between new and existing content, observing whether participants asked the AI to generate entirely new material, extend existing material, or refine previously written text. The latter two are further qualified by labeling whether the text being extended or refined was generated by the user or the agent.

Following the coding process, I conducted quantitative analyses of the dialogue patterns, including frequency distributions of primary dialogue acts, request types, directness, constraint, focus areas, and content operations. I created participant-level summaries calculating the percentage of each dialogue characteristic for individual participants, developed a dialogue pattern typology that categorized participants based on their dominant request type (e.g., “Primarily Content Generation”) and constraint pattern (e.g., “Primarily Bounded”), and performed cluster analysis using dialogue act counts to identify natural groupings of interaction patterns.

6.5.2.2 Results

Requests dominate the interactions (96 out of 101 messages, 95%), with the remaining messages consisting of clarifications/elaborations (2%), assessments (1%), and other communications (2%). This overwhelming prevalence of requests suggests that participants primarily adopted an instrumental stance toward the AI, positioning it as a tool to accomplish specific tasks rather than as a collaborative partner.

Content generation directives represent the majority of requests (74% of all requests), highlighting that participants primarily engaged with the AI as a content producer. Information-seeking directives (13.5%) and evaluation requests (5.2%) form a smaller but notable portion of interactions, indicating secondary roles for the AI as an information source and evaluator.

Most requests (82.3%) are direct rather than indirect dialogue acts. Similarly, bounded requests (75%) are three times more common than unbounded requests (25%), suggesting that participants typically provided specific constraints or parameters rather than giving the AI complete creative freedom.

Plot-related requests dominate the focus areas (57.3%), followed by stylistic concerns (22.9%) and structural elements (16.7%). This emphasis on plot development reflects the narrative nature of the task, with participants primarily seeking assistance with story progression rather than linguistic or organizational aspects. Among content generation requests, extending existing content was the most common operation (31 instances), followed by generating new content and refining existing content. This suggests that participants frequently built upon established narrative elements rather than continuously introducing new ideas or refining existing ones.

The dialogue pattern typology analysis revealed distinct interaction strategies across attribution quadrants (see Section 6.5.1). High Delegation participants (9) exclusively used content generation-focused strategies, with 5 classified as “Primarily Content Generation - Primarily Bounded.” Similarly, participants in the “Expressive Delegation” quadrant (2) also focused solely on content generation. In contrast, High Agency participants (12) showed much greater diversity in strategies, including information seeking (3), evaluation (1), and mixed approaches (4). This pattern was further supported by the Kruskal-Wallis test, which showed highly significant differences in content generation counts across attribution quadrants ($\chi^2 = 15.36, p < 0.001$). Figure 6-9 presents these counts as a heatmap.

The cluster analysis identified two natural groupings of dialogue patterns: a larger cluster (20 participants) with lower content generation (avg 2.05) and some information seeking (avg

0.95), and a smaller cluster (3 participants) with much higher content generation (avg 10) and minimal information seeking (avg 0.33). The second cluster contained primarily High Delegation participants (2 out of 3), suggesting a distinct “heavy AI delegation” interaction pattern.

Logistic regression confirmed that content generation was the strongest predictor of High Delegation attribution. Each additional content generation request significantly increased the odds of being in the High Delegation quadrant ($\beta = 0.61$, $p = 0.017$). Unbounded requests showed a marginally significant effect ($\beta = 0.88$, $p = 0.052$), providing some evidence that less constrained requests may contribute to greater AI contribution.

Ratio analysis, which addresses the mutual exclusivity constraint of request types, showed that the ratio of content generation to total requests was much higher in High Delegation (0.982) and Expressive Delegation (1.0) compared to High Agency (0.486). Conversely, the ratio of information seeking to total requests was much higher in High Agency (0.375) compared to High Delegation (0.008). These findings suggest that different dialogue strategies strongly correspond to different attribution outcomes. Based on these analyses, three primary interaction strategies emerge:

Bounded Content Generation: Characterized by high constraint (bounded requests) combined with content generation, this approach allowed participants to maintain creative direction while leveraging the AI’s text generation capabilities. These participants often provided specific plot points, character details, or stylistic guidelines, maintaining creative control through constraints.

Unbounded Content Generation: Characterized by low constraint (unbounded directives) combined with content generation, this approach gave the AI greater creative latitude, often resulting in higher AI contribution to both ideational and textual dimensions. Participants using this strategy typically provided minimal direction and incorporated large portions of AI-generated content.

Information-Seeking: Characterized by requests for information rather than content generation, this approach positioned the AI as an advisor rather than a co-creator. Participants

using this strategy maintained almost complete control over text production (averaging 93.5% participant-authored text) while consulting the AI for ideational input or feedback.

6.5.3 RQ2: Questionnaires

This analysis examines participants' subjective experiences and attitudes during the activities and on creativity and writing generally. By examining post-activity evaluations, creativity measures, writing attitudes, and AI experience, this section provides insight into how participants perceived the differences between AI-assisted and individual writing processes, as well as their creative self-efficacy, creative personal identity, writing attitudes, writing self-efficacy, and their prior experience with different large language models.

6.5.3.1 Post-activity evaluation

After each writing activity participants completed a questionnaire about their writing experience (Table F). Participants rated eight dimensions (fun, enjoyment, expressiveness for creativity and self-expression, outcome satisfaction, ease of writing for starting and finishing, and ownership) on a five-point Likert scale from “Strongly Disagree” to “Strongly Agree.” After completing both activities, participants directly compared their experiences between the individual and AI-assisted conditions by selecting their preferred method for each of these same dimensions, allowing for direct comparison between the two writing approaches. These questionnaires were designed to quantify the co-creative writing experience [82]. I calculated scores by converting the five-point Likert scale responses to numerical values (1=“Strongly Disagree” to 5=“Strongly Agree”).

Post-activity evaluations revealed interesting patterns in participants' experiences with both writing conditions. Figure 6-10 illustrates the distribution of Likert responses across all eight metrics, with several notable observations.

The most striking difference appears in the ownership metric, where participants rated individual writing ($M = 4.68$, $SD = 0.95$) higher than AI-assisted writing ($M = 3.75$, $SD = 1.52$). This pattern was reinforced in the preference assessment (Figure 6-11), where 14

participants selected individual writing as producing stories they felt were “most theirs,” compared to only 6 selecting AI-assisted writing.

For ease of writing, the distributions suggest an advantage for AI assistance. Both ease of starting (AI-assisted $M = 4.35$, $SD = 0.99$; Individual $M = 3.65$, $SD = 1.14$) and ease of finishing writing (AI-assisted $M = 3.95$, $SD = 1.05$; Individual $M = 3.60$, $SD = 1.23$) show higher ratings for the AI-assisted condition. The preference questions also reflect this pattern, with 13 participants selecting AI-assisted writing as easier for both starting and finishing, compared to only 6 selecting individual writing.⁷

For the *enjoyment* measure, the AI-assisted condition had a slightly higher mean ($M = 4.15$, $SD = 1.14$) than the individual condition ($M = 3.95$, $SD = 1.00$). Additionally, when asked to select a preference between the two conditions, 12 participants selected AI-assisted writing as their preferred method for future short story writing, compared to 8 selecting individual writing.

The remaining metrics show similar distributions between conditions. The *fun* metric showed nearly identical means (AI-assisted $M = 4.5$, $SD = 0.76$; Individual $M = 4.3$, $SD = 0.80$) and an exact split in preference (10 vs. 10). Similarly, *outcome satisfaction* show minimal differences (AI-assisted $M = 3.7$, $SD = 0.98$; Individual $M = 3.75$, $SD = 0.85$), though the preference lean toward AI-assisted (12 vs. 8). Likert ratings for both *creativity* (AI-assisted $M = 4.15$, $SD = 0.75$; Individual $M = 4.15$, $SD = 1.04$) and *self-expression* (AI-assisted $M = 3.9$, $SD = 0.97$; Individual $M = 3.85$, $SD = 1.09$) show similar mean scores across conditions. Additionally, participants strongly favored individual writing for *self-expression* (13 vs. 6), a preference ratio that mirrors the *ease of writing* metrics (6 vs. 13) with much smaller differences in the corresponding Likert ratings.

6.5.3.2 Creative self-efficacy and creative personal identity

After completing the post-activity preferences evaluation, participants completed a questionnaire designed to measure their own perceptions of their creative abilities and the importance of creativity to their identity[86]. This 11-item assessment included five statements

⁷ One participant (ID 74) did not answer four of the eight preference questions, including both ease of writing measures and both expressiveness measures.

specifically targeting creative personal identity (e.g., “Creativity is an important part of myself”) and six statements assessing creative self-efficacy (e.g., “I trust my creative abilities”).

Participants rated their agreement with each statement on a scale ranging from “Definitely Not” to “Definitely Yes,” providing insight into how participants viewed their creative capabilities and how central creativity was to their self-concept (Table G). I computed separate subscale scores by averaging responses across the respective items (five for creative personal identity and six for creative self-efficacy). Responses were converted to numerical values on a five-point scale, with higher scores indicating stronger creative identity and greater confidence in creative abilities.

Results: Participants reported moderately high levels of creative identity and efficacy. Creative Personal Identity (CPI) scores averaged 4.04 out of 5 ($SD = 0.97$), with a median of 4.20, indicating that most participants viewed creativity as an important aspect of their self-concept. The distribution showed some variation, with scores ranging from 1.80 to 5.00. Similarly, Creative Self-Efficacy (CSE) scores were generally positive ($M = 4.18$, $SD = 0.67$), with a median of 4.50 and a range from 2.67 to 5.00. These results suggest that participants generally trusted their creative abilities and identified with creativity as a personal characteristic. The slightly higher mean and lower standard deviation for CSE compared to CPI indicates that participants were somewhat more confident in their creative abilities than they identified with creativity as a core aspect of their identity. Notably, 25% of participants rated their creative self-efficacy at 4.67 or higher, demonstrating substantial confidence in their creative capabilities.

6.5.3.3 Writing attitudes and writing self-efficacy:

Participants were also asked to complete a questionnaire designed to evaluate their own feelings toward writing and confidence in their writing abilities [63, 159]. The questionnaire consisted of 11 statements, with six items measuring *writing attitudes* (e.g., “I usually enjoy writing”) and five items assessing *writing self-efficacy* (e.g., “I can put my ideas into writing”). Participants rated each statement from “Definitely Not” to “Definitely Yes,” providing insight into participants’ relationships with writing and perceived competence (Table H). I computed separate subscale scores by averaging responses across the respective items (six for writing attitudes and

five for writing self-efficacy). Responses were converted to numerical values on a five-point scale, with higher scores indicating more favorable attitudes towards writing and greater confidence in writing abilities. When computing the writing attitudes score, I reverse-scored the negative statement “I try to avoid writing as much as possible.”

Results: Writing-specific measures revealed moderate positive attitudes and self-efficacy. Writing Attitudes (WA) scores averaged 3.69 out of 5 ($SD = 0.89$), with a median of 3.92, suggesting generally positive but varied attitudes toward writing across participants. Writing Self-Efficacy (WSE) showed a similar pattern ($M = 3.63$, $SD = 0.86$), with a median of 3.70. Both measures displayed considerable variation, with WA ranging from 1.83 to 5.00 and WSE ranging from 1.80 to 5.00. Interestingly, the first quartile for both writing measures ($WA = 2.96$, $WSE = 2.95$) was noticeably lower than for the creativity measures, indicating that a quarter of participants had somewhat negative attitudes toward writing or doubted their writing abilities. The data also revealed that participants generally rated their writing attitudes slightly higher than their writing self-efficacy, suggesting that some participants may enjoy or value writing despite feeling less confident in their writing abilities. Only two participants (10%) achieved the maximum score of 5.00 on both writing measures, compared to three participants (15%) who scored at or below 2.60 on both measures.

6.5.3.4 AI experience

Participants were also asked to rate their familiarity with and usage patterns of contemporary large language models. Unlike the original instrument [185], which focused exclusively on ChatGPT, this adapted version expanded coverage to include five prominent AI systems available at the time of the study: ChatGPT, Claude, Gemini, Grok, and an option for other AI language models not listed. For each system, participants indicated their usage frequency on a seven-point scale ranging from complete unfamiliarity (“I do not know this model”) to frequent engagement (“Daily use”). DeepSeek was notably absent from this assessment as its public release coincided with the beginning of study sessions, making questionnaire modifications impractical at that stage.

Results: Analysis of participants’ prior experience with AI language models, shown in Figure 6-12, revealed varying levels of familiarity across different systems. ChatGPT was the most widely used model, with all participants having at least some experience: 30% reported daily usage, 40% used it several times weekly, and the remaining 30% had occasional experience. Claude showed moderate familiarity, with 25% of participants using it weekly or daily, though 25% were unfamiliar with the model entirely.

Gemini had higher recognition but lower regular usage than Claude—all participants had at least heard of the model, but 60% had never used it. Only one participant reported using Gemini several times a month, which was the highest usage frequency reported for this model. Grok was the least recognized and had the least usage, with 55% of participants reporting no knowledge of it and only one participant having used it previously. The remaining 40% had heard of Grok but never used it. When asked about “other” AI language models, 70% of participants either did not know or had only heard of alternative models, indicating that participants’ AI experience was largely concentrated among the major commercial platforms, with ChatGPT being the predominant system.

6.5.3.5 Language background

At end of the study, participants were asked to complete a questionnaire about their language background, which were adapted from the Language and Social Background Questionnaire [6]. It first established the number of languages spoken. For multilingual participants, it identified whether English was the participant’s first language. If not, participants were asked to provide their first language. The participants with English as an additional language were also asked about language exposure across four developmental periods (infancy, preschool, primary school, and high school) and current language usage patterns across four communication modalities (speaking, listening, reading, and writing). This information provided context for understanding potential linguistic factors that might influence participants’ writing processes and their interaction with AI tools.

During the oral consent process and study introduction, I informed participants that their questionnaire responses were valuable for my research, but emphasized their right to skip any questions they preferred not to answer, without risk to their research participation credit. I explicitly stated that I would rather participants leave questions blank if they felt uncomfortable or uncertain (or even just disinclined) to reduce pressure to provide random or false answers simply to complete the study requirements.

To ensure data quality, I incorporated six attention check questions arranged throughout the questionnaires. These checks employed various validation methods, following current research recommendations on detecting inattentive responding [42]. Based on attention check performance, three participants with multiple failures were excluded from analysis. A detailed description of the attention check methodology, scoring system, and exclusion criteria is provided in Appendix K.

6.6 Discussion

The overarching aim of this study is to investigate how conversational AI influences both the narrative writing process and the resulting narrative artifacts. This findings of this study provide insight into divergent patterns of creative task agency and delegation and balancing creative control with technological assistance. During the AI-assisted writing condition, some participants maintain creative control over both ideation and authorship, while others delegate extensively to the AI agent across both dimensions.

This discussion examines these patterns and their implications through a central narrative: the balancing of creative agency between human and AI in collaborative writing. First, I present key observations about the distribution of creative agency between human writers and AI, analyzing the emerging typography of collaboration strategies and their relationship to creative ownership. Second, I investigate how dialogue patterns serve as mechanisms through which this agency is negotiated, revealing how different approaches to AI engagement correspond to different distributions of creative responsibility. Third, I discuss the linguistic and semantic consequences

of these agency decisions, identifying a fundamental tension between individual enhancement and collective homogenization.

Throughout this discussion, I integrate quantitative findings with qualitative insights to construct an understanding of how people incorporate AI into their creative processes.

6.6.1 Creative Agency and Ownership in Human–AI Collaboration

6.6.1.1 Creative agency-delegation typography

The two-dimension content attribution analysis reveals distinct patterns in how participants distributed creative responsibility between themselves and the AI agent. As illustrated in Figure 6-7, participants cluster primarily in either the High Agency quadrant (52.2%) where they maintained control over both ideation and textual expression, or the High Delegation quadrant (9), where they ceded all textual expression and most, or all, ideational responsibilities to the agent. A small minority (2) are classified into the Expressive Delegation quadrant, where they contributed more of their own ideas, but still relied heavily on AI-generated text.

This polarization suggests that writers may conceptualize AI collaboration in binary terms—either as a tool under their direction or as an independent creator to whom they delegate responsibility—rather than as a fluid partnership with shared creative control. The clustering in opposing quadrants suggests that people tend toward either maintaining high creative ownership or substantially delegating the creative process, with few adopting hybrid approaches.

Moreover, the complete absence of participants in the Conceptual Delegation quadrant, which represents using predominantly AI-originated ideas expressed through participant-authored text, reveals a fundamental asymmetry in how creative control is valued during human-AI collaboration, suggesting that more value is placed on the conceptual direction than the technical execution.

Figure 6-8 provides further insights into how participants integrated AI contributions to their narratives. High Agency participants typically incorporate AI contributions as discrete, bounded segments within predominantly self-authored narratives. This selective integration suggests an instrumental view of AI as an occasional advisor or assistant rather than a true

collaborator. In contrast, High Delegation participants produced narratives with large, continuous blocks of AI-authored text, often with minimal intervention between requests. This pattern suggests these participants may view their creative role fundamentally differently—perhaps as directors who specify a vision but delegate execution or as curators who select and approve content rather than generating it directly.

Interestingly, despite having fewer interactions with the agent on overall, the High Agency quadrant displays greater variation in content integration compared to the High Delegation quadrant. Some High Agency participants incorporate small segments of AI-generated text or AI-originated ideas throughout their narratives, while others used AI contributions only at specific points in the narrative development. This variation indicates that even among those who maintain high creative agency, there are multiple strategies for integrating AI contributions.

6.6.1.2 Dialogue strategies as mechanisms for managing creative agency

The dialogue act analysis reveals how participants' interaction strategies correspond to different patterns in the Agency-Delegation Typology. The overwhelming prevalence of request dialogue acts (95% of all interactions) suggests that participants primarily approached the AI as a tool to accomplish specific tasks rather than as a collaborator or conversational partner. Within this framework, three distinct interaction strategies emerge from the analysis, discussed in [Section 6.5.2](#), each corresponding to different patterns of creative control: Bounded Content Generation, Unbounded Content Generation, and Information-Seeking.

The contrast between content generation and information-seeking strategies reveals fundamentally different conceptualizations of the AI's role in the creative process. Participants who primarily used content generation requests position the AI as a co-creator or primary content producer, delegating substantial creative responsibility. In contrast, those who primarily made information-seeking requests position the AI as an advisor or resource while maintaining primary creative control themselves. This distinction demonstrates how dialogue strategy directly shapes the AI's functional role within the creative partnership.

Furthermore, the significant differences in dialogue diversity across attribution quadrants reveal how interaction strategy shapes creative outcomes. High Agency participants employed a much wider range of dialogue strategies, including information seeking, evaluation, and mixed approaches. This diversity suggests a more flexible, selective engagement with AI capabilities. In contrast, High Delegation participants almost exclusively used content generation strategies, with most adopting primarily bounded approaches. This pattern indicates that dialogue diversity itself may be a mechanism for maintaining creative control—by engaging with the AI across multiple dimensions, participants maintain greater agency in the creative process.

The ratio analysis provided particularly compelling evidence of this relationship—the ratio of content generation to total requests was much higher in High Delegation (0.982) and Expressive Delegation (1.0) compared to High Agency (0.486). Conversely, the ratio of information seeking to total requests was much higher in High Agency (0.375) compared to High Delegation (0.008). These findings demonstrate that dialogue strategy strongly predicts creative attribution outcomes, establishing a clear connection between interaction patterns and agency distribution.

6.6.2 Enhancement and Homogenization

The linguistic analyses revealed a trade-off in AI-assisted creative writing: while AI assistance enhances certain aspects of individual texts, it simultaneously appears to constrain the diversity of expression across writers. This trade-off represents an important aspect of understanding the impact of AI assistance on narrative artifacts.

At the most basic level, AI-assisted texts were significantly longer than individually written texts, with a medium-to-large effect size. However, the substantial variance in the AI-assisted condition (variance ratio exceeding 19) reveals that AI's impact on text production varies dramatically across individuals. While some participants leveraged AI to produce substantially longer narratives, others showed more modest increases. This variability, shown in Figure 6-13, suggests that AI assistance does not deterministically increase production but rather amplifies individual differences in how writers engage with the technology.

Beyond simple text length, the significant increases in lexical diversity (MSTTR) and sophistication (LS1) (Section 6.4.2) in AI-assisted texts suggest that AI collaboration may enhance certain dimensions of linguistic quality. Participants working with AI produced texts with more varied vocabulary and more sophisticated word choices. However, no evidence of changes in syntactic features emerge in measures of sentence complexity, coordination, or subordination.

Additionally, analysis of referential cohesion (Section 6.4.3) indicated a significant interaction between condition and text length. In individual writing, referential cohesion decreased as text length increased. In contrast, AI-assisted texts maintained more consistent, but generally lower, referential cohesion regardless of length, and may flatten variations in cohesion as text length increases.

Perhaps the most interesting finding is the significant increase in semantic homogeneity among AI-assisted narratives, despite their greater lexical diversity. While AI-assistance may enhance the linguistic sophistication of individual narratives, it appears to constrain the diversity of conceptual approaches across different writers.

The hierarchical clustering dendrograms (Figure 6-14) visualize this homogenization effect, with AI-assisted narratives forming more compact clusters with shorter branch heights. These visualizations reveal the substantial differences in semantic homogeneity between conditions. The AI-Assisted condition (left) demonstrates more compact clustering with shorter branch heights, indicating stronger semantic similarities between narratives. Particularly notable is the formation of two distinct clusters (shown in blue and orange), suggesting that AI-assisted narratives converged toward common semantic patterns. In contrast, the Individual condition (right) displays greater branch heights and more varied clustering patterns, with multiple smaller clusters of different colors representing greater semantic diversity. Additionally, the dendrogram branches in the Individual condition extend higher up the y-axis, indicating greater semantic distances between narratives. This finding raises important questions about AI's influence on creative diversity: *does AI assistance, while enhancing certain aspects of individual texts, ultimately narrow the range of creative expressions across writers?*

When considered in relation to the agency distribution patterns discussed earlier, these linguistic findings suggest a deeper connection between creative agency and textual outcomes. The homogenization effect appears to be an emergent property of distributed creative agency: As participants delegate more creative responsibility to the AI, their narratives converge toward common semantic patterns despite surface-level linguistic diversity. This tension between individual enhancement and collective homogenization may be an important consideration for educational contexts, creative writing communities, and professional content creation, where maintaining diverse perspectives and approaches is often valued.

6.6.3 Implications for Human-AI Co-Creativity

The polarized distribution of creative agency suggests that current AI writing tools may encourage binary rather than nuanced collaborative relationships. This has implications for how we design, teach about, and evaluate AI writing assistants. Future interfaces might need to explicitly support intermediate collaboration patterns that balance human and AI contributions, while educational approaches should help users recognize and explore the full spectrum of possible collaborative relationships rather than defaulting to either minimal or maximal delegation.

The strong relationship between dialogue strategies and creative outcomes suggests that interaction patterns could serve as early indicators of how users will distribute creative responsibility. This could inform the development of adaptive systems that recognize and respond to different collaboration styles. For example, systems could identify when users are employing primarily content generation requests and offer alternative interaction strategies to help them maintain creative ownership if desired, or provide tailored support based on detected dialogue patterns.

The tension between individual text enhancement and collective homogenization suggests a fundamental challenge for AI-assisted creativity. Educational institutions, creative industries, and AI developers must consider whether the linguistic benefits of AI assistance outweigh potential constraints on creative diversity. This trade-off may necessitate new approaches to diversifying AI outputs, such as personalization techniques that specifically optimize for uniqueness across users,

or educational practices that emphasize maintaining distinctive voice while leveraging AI capabilities.

The absence of Conceptual Delegation patterns indicates that users may value idea ownership more highly than textual execution when collaborating with AI. This hierarchy of creative control could inform future models of creative attribution and responsibility in human-AI collaboration. For instance, attribution frameworks might need to distinguish between conceptual and textual contributions, while educational assessment may need to place greater emphasis on evaluating students' conceptual input rather than their textual execution. Similarly, AI writing interfaces could be designed to better support humans in maintaining conceptual control while offering textual assistance.

The significant differences in linguistic features between AI-assisted and individual writing raise important questions for educational assessment. Traditional writing evaluation metrics may need reconsideration in contexts where AI assistance is available. Educators might need to develop new assessment approaches that focus on the writing process rather than just the final product, evaluate students' critical engagement with AI suggestions, or place greater emphasis on aspects of writing that remain distinctly human despite AI assistance, such as the development of a personal voice or the integration of lived experience into narrative.

6.7 Limitations and Future Directions

While this study provides valuable insights into patterns of human-AI co-creativity, several limitations suggest directions for future research. The sample size ($n=23$) and demographic homogeneity (university students) limit generalizability, suggesting the need for larger-scale investigations across diverse populations. The laboratory setting, while enabling controlled comparison, may have influenced natural writing behaviors; future studies in more naturalistic contexts could validate these patterns.

Additionally, this study captured a snapshot of co-creative writing at a specific technological moment. As AI capabilities evolve, so too may patterns of human-AI creative collaboration. Longitudinal studies could track how these patterns shift over time as users gain familiarity with

AI tools and as the tools themselves advance, providing insight into the dynamic evolution of human-AI creative relationships.

Future research might also explore how different prompting or interface designs influence the distribution of creative agency, how collaboration patterns vary across different creative domains beyond narrative writing, and how individual differences in creative identity and writing experience moderate engagement with AI assistance. These investigations would further enrich our understanding of the complex interplay between human creativity and AI capability.

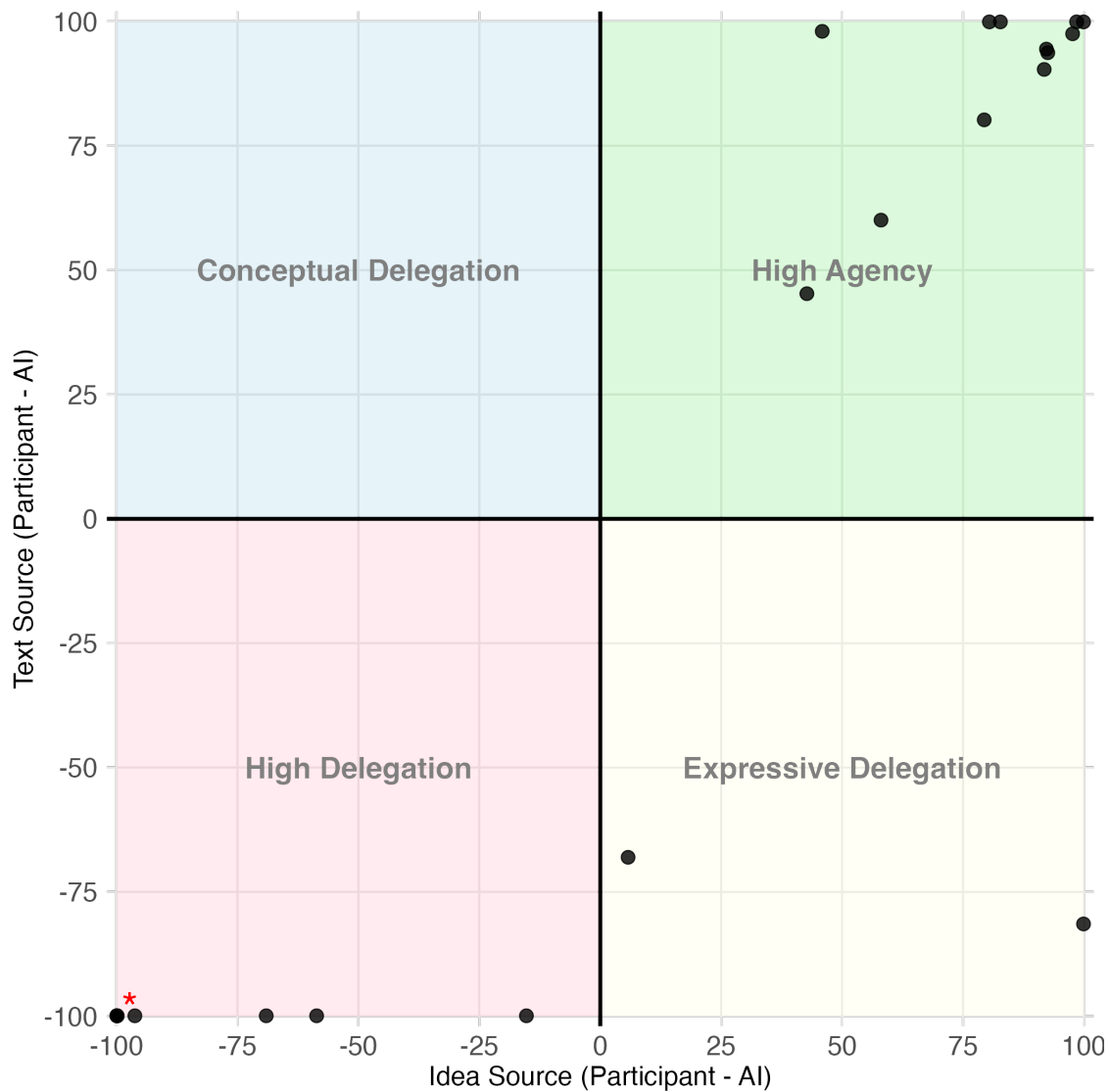


Figure 6-7. Co-creative attribution mapping of participants' artifacts (n=23) across ideational and textual source dimensions. The x-axis represents the distribution between participant-originated and AI-originated ideas; the y-axis represents participant-authored versus AI-authored text. Quadrants indicate distinct co-creative patterns: High Agency (top right), High Delegation (bottom left), Expressive Delegation (bottom right), Conceptual Delegation (top left). Asterisks indicate multiple participants at the same position.

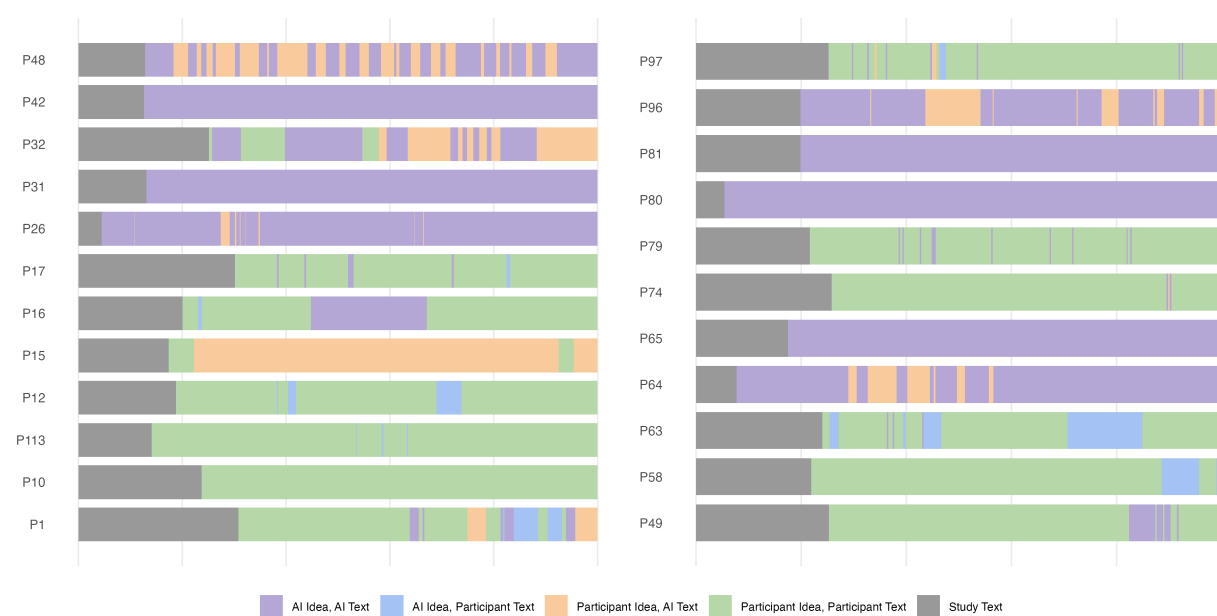


Figure 6-8. Distribution and sequence of different attribution sources across participant documents. Each segment's width represents its proportional percentage of the total document text.

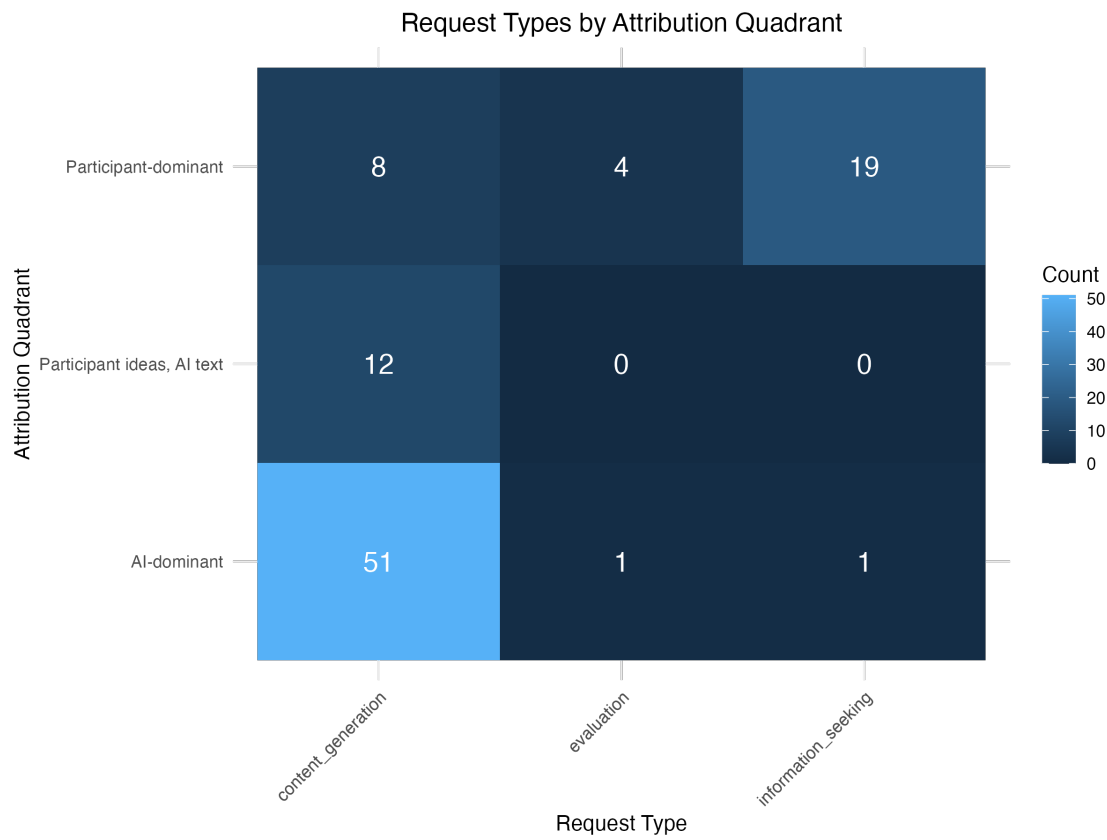


Figure 6-9. Heatmap showing the distribution of request types across attribution quadrants. Cell values indicate the raw count of each request type.

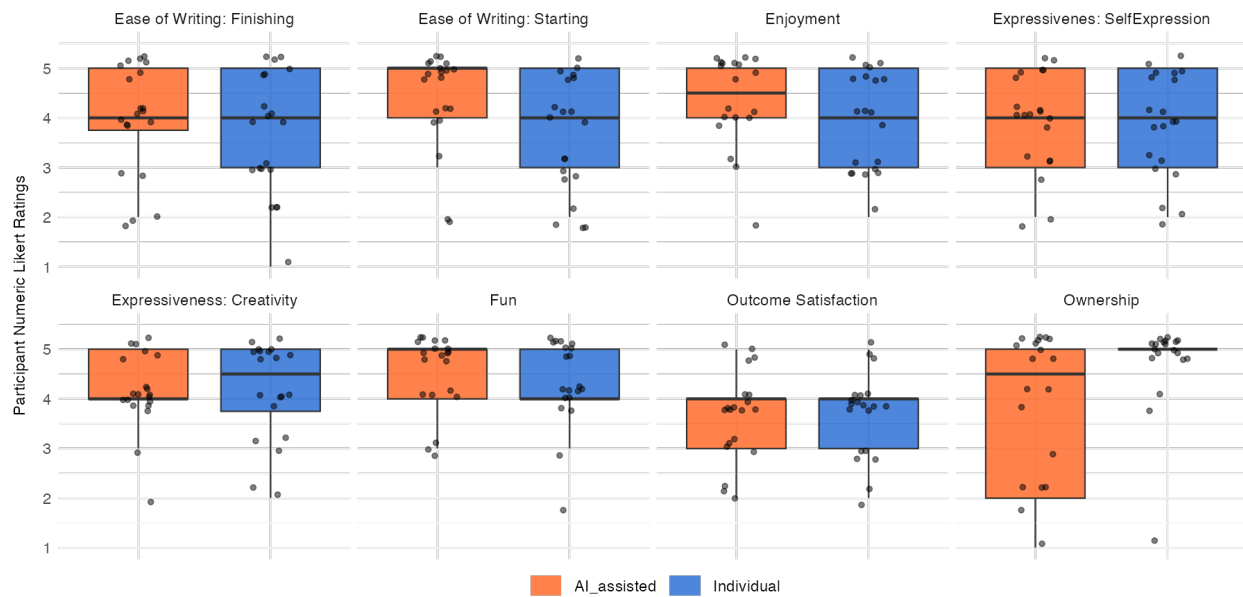


Figure 6-10. Post-Activity Likert Responses by Writing Condition. Box plots showing the distribution of participant ratings (1-5 scale, 1=Strongly Disagree, 5=Strongly Agree) across eight writing experience metrics, comparing AI-assisted (orange) and individual (blue) writing conditions.

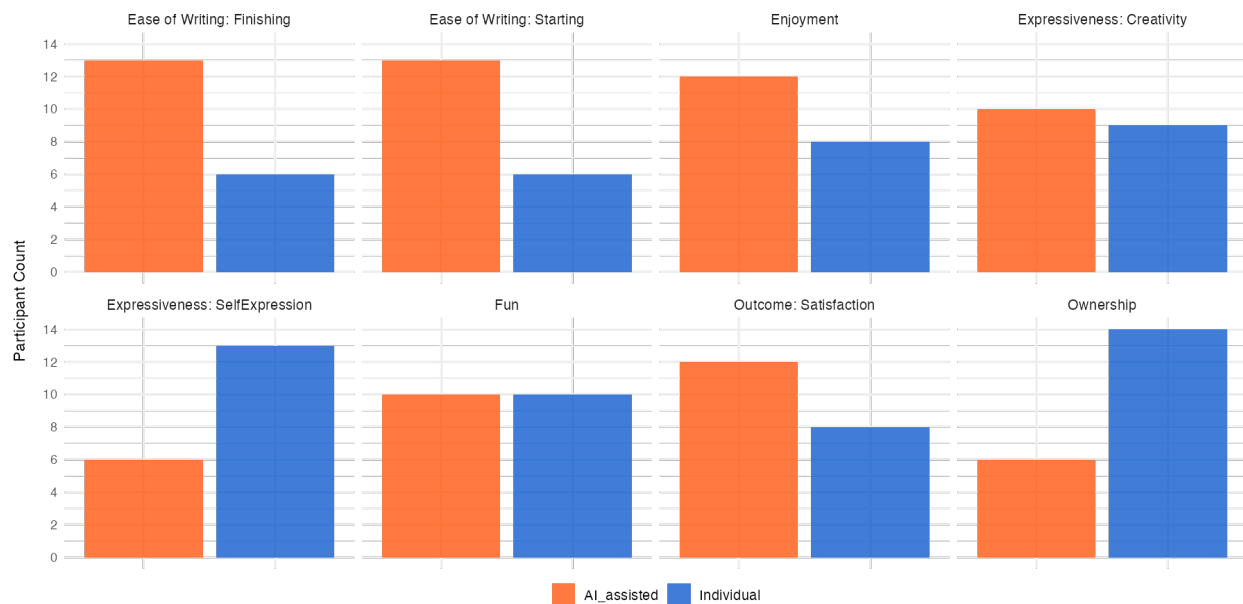


Figure 6-11. Bar chart showing count of participants' preferences when required to select either the AI-assisted (orange) or individual (blue) writing condition for each of eight writing experience metrics.

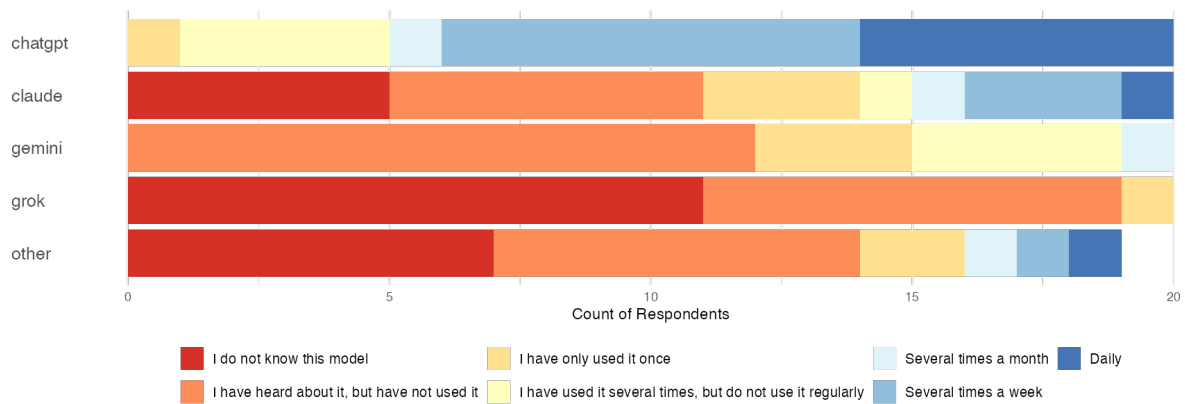


Figure 6-12. Distribution of experience levels across five AI systems. Experience levels range from “I do not know this model” (leftmost) to “Daily” usage (rightmost).

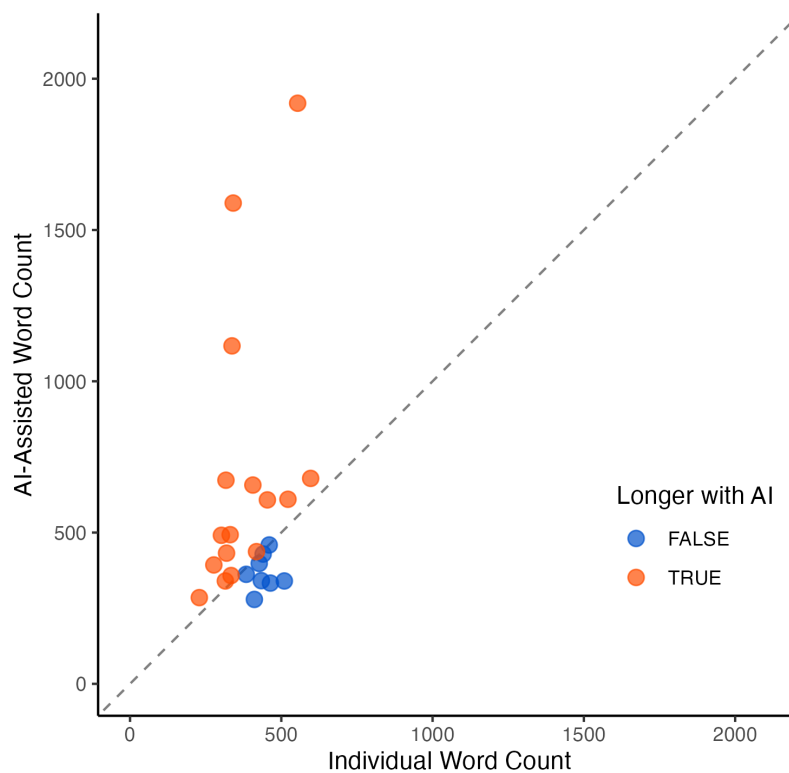


Figure 6-13. Text lengths comparing AI-Assisted and Individual Writing. Points above the diagonal line indicate longer text with AI assistance.

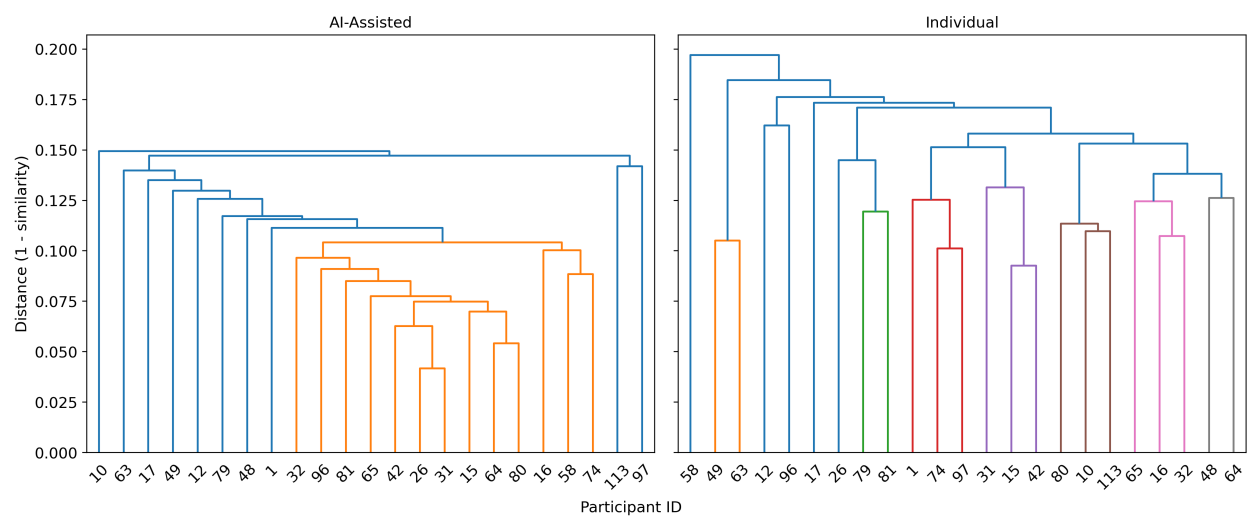


Figure 6-14. Hierarchical clustering dendrograms visualizing semantic similarities between narratives. The y-axis represents distance (1 - similarity), with lower values indicating higher similarity. Left: AI-Assisted condition shows tighter clustering with smaller distances between stories, forming more distinct groups (orange and blue clusters). Right: Individual condition exhibits greater semantic diversity with larger distances between clusters and more varied branching patterns.

CHAPTER 7 CONCLUSION

This dissertation investigates human-AI co-creativity through a series of studies examining collaboration in different creative domains. The research addresses the following questions:

1. What are the dialogue states that characterize the co-creative process between learners, and how do learners transition between these dialogue states?
2. What are the relationships between the outcome of partner satisfaction and the dialogue that comprise the co-creative process?
3. What challenges and opportunities emerge when designing and evaluating a co-creative AI for an expressive CS learning environment?
4. How does conversational AI assistance impact the users' final narrative artifacts in terms of text length, linguistic complexity, and cohesion compared to narratives written individually?
5. How does conversational AI assistance influence users' creative decision-making and narrative development?

This chapter summarizes the research progression, synthesizes key findings across the studies, discusses limitations and future research directions, and offers concluding remarks.

7.1 Summary of Research Progression and Findings

This section outlines the core findings corresponding to the progression of the research presented in the preceding chapters.

7.1.1 Studies 1 & 2: Understanding Co-Creativity

This dissertation establishes a foundational understanding of co-creativity through analysis of human-human co-creativity within a computational music remixing context. In the work reported in Chapter 3, I employ a hidden Markov model and identify seven distinct dialogue states that characterize the co-creative process between learners in a computational music remixing environment: Social Dialogue, Aesthetic Dialogue, Curriculum Browsing, Code Editing,

Technical Dialogue, Compilation Success, and Compilation Error. This analysis reveals how learners transition between these states during co-creative work, with most sessions beginning in Social Dialogue before moving through various combinations of creative and technical states.

The identification of these dialogue states provides insights into the structure of co-creative interactions. Particularly significant is the distinction between Aesthetic Dialogue—characterized by creative proposals and acceptances—and Technical Dialogue—focused on code features and task requirements. This distinction highlights the dual nature of co-creativity, which requires navigating both creative expression and technical implementation.

Building on this foundation, Chapter 4 examines the relationship between these dialogue states and partner satisfaction, finding that sustained dialogue, particularly during challenges, correlates with higher partner satisfaction, whereas specific state transitions (e.g., Aesthetic to Technical Dialogue) and reduced dialogue post-transition to coding are associated with lower partner satisfaction. These studies characterize co-creative dialogue patterns and linked them to collaborative outcomes.

7.1.2 Study 3: Design and Evaluation of a Co-Creative AI Agent

Informed by the findings on human-human collaboration, Chapter 5 details the iterative design and evaluation of EarSketch CAI, a co-creative AI (CAI) agent intended to support students learning computational music creation. We designed CAI to support both the aesthetic and technical facets of the co-creative process. The study reveals divergent user expectations: some users primarily seek creative assistance (sound suggestions), while others primarily seek debugging support. This variation in expectations parallels the division of labor between the aesthetic and technical aspects of the projects observed in human-human collaboration.

Additionally, the findings reveal adoption challenges that are absent in human-human collaboration, suggesting that students need time to understand, explore, and feel comfortable with the new technology.

Finally, the summer camp interviews reveal how students' experiences with other AI agents, particularly large language models like ChatGPT, shape their expectations of other AI agents.

This finding highlights the importance of considering users' broader technological context when designing co-creative AI systems.

7.1.3 Study 4: Impact of AI Assistance on Narrative Writing

Building upon the insights gained from studying human-human dialogue and designing the CAI agent, Chapter 6 investigates the effects of conversational AI assistance on narrative writing. Analysis of the resulting artifacts reveals that AI assistance leads to significantly longer texts with increased lexical diversity (MSTTR) and sophistication (LS1), but also significantly higher semantic similarity across narratives compared to unaided writing.

Additionally, the analysis of the creative process reveals a polarization of creative agency: participants tended towards either High Agency (maintaining control) or High Delegation (ceding control to the AI), with few adopting intermediate approaches. Dialogue analysis indicates that interaction strategies predict agency distribution; High Agency users employ more diverse strategies, while High Delegation users predominantly make content generation requests.

These findings suggest that AI assistance does not deterministically shape creative outcomes but rather reflects individual differences in how people approach creative collaboration. The substantial variation in agency distribution, dialogue strategies, and resulting artifacts suggests that understanding these individual differences is important for designing effective co-creative systems.

7.2 Dynamics of Co-Creative Partnership

The findings across these studies, spanning different domains and methodologies, reveal key dynamics of human-AI creative partnerships.

7.2.1 The Centrality of Dialogue in Collaboration

A central finding is that co-creativity, regardless of the partners' nature (human or AI), is fundamentally rooted in dialogue. The analysis of human-human music remixing (Chapter 3) identified distinct dialogue states (e.g., Social, Aesthetic, and Technical Dialogue) for navigating the co-creative process. These states reveal the dynamic shifts between creative exploration and technical execution. This understanding directly informs the design of the EarSketch CAI system

(Chapter 5), which supports both the creative and technical facets of computational music creation through targeted dialogue modules.

The importance of dialogue patterns persisted in the human-AI narrative writing context (Chapter 6). Interaction strategies significantly correlate with how creative agency was distributed. Participants maintaining high control over the artifact (High Agency) utilize a diverse range of *request* dialogue acts. Conversely, those delegating significant responsibility (High Delegation) rely almost exclusively on content generation requests. The stark difference in the content generation request ratio between High Delegation (0.982) and High Agency (0.486) underscores how dialogue strategies reflect and shape the collaborative dynamic and agency distribution.

These findings demonstrate a consistent relationship between interaction patterns and co-creative outcomes across different domains and collaborator types.

7.2.2 Agency, Asymmetry, and Creative Contribution in Human-AI Interaction

While dialogue forms a common foundation, a critical distinction arises in the asymmetry inherent in current human-AI collaborations compared to human-human partnerships. Unlike human collaborators who share agency and intrinsic motivation, AI systems currently lack genuine ownership or creative intent. This asymmetry shapes human engagement with AI partners. In the narrative writing study (Chapter 6), the AI's lack of agency allowed users to fully control the distribution of responsibility, leading to the observed polarization (High Agency vs. High Delegation). This finding contrasts with the more balanced negotiated agency dynamics typical of human-human collaboration (Chapter 4).

The research indicates that users conceptualize and interact with AI creative partners in different ways, often adopting relatively simplistic or binary views of the AI's utility. The formative evaluation of CAI (Chapter 5) shows users seek one of two distinct types of help—aesthetic or technical. This parallels the findings in narrative writing (Chapter 6), where dialogue strategies reflect primarily two views of the AI: as a content producer (delegation via generation requests) or as an advisor/resource (agency via information-seeking/evaluation), neglecting a wider spectrum of potential collaborative interactions.

Additionally, the prevalence of *request* dialogue acts (Chapter 6) further suggests that users, even when delegating, primarily interacted with the AI as a means to an end, rather than a peer collaborator.

7.3 Open Questions and Future Research Directions

This dissertation, while providing significant insights, raises fundamental questions and highlights promising avenues for future research into human-AI co-creativity.

7.3.1 Creative Identity and AI Engagement

The polarization in agency distribution observed in the narrative writing study and the divergent support-type expectations of learners during the CAI evaluations suggests that individual differences shape how people engage with AI creative tools. These findings raise new questions: *What aspects of creative identity, experience, or self-efficacy predict different patterns of engagement?* and *How might these factors be addressed in the design of co-creative systems?*

Future research should investigate how factors such as creative identity, self-efficacy, prior experience, and creative goals influence patterns of AI engagement. Understanding these individual differences could enable more personalized approaches to co-creative system design and more targeted interventions to support beneficial integration of AI into creative practices. It could also identify which individuals might be most susceptible to over-reliance on AI assistance and inform strategies to mitigate this risk.

7.3.2 Evolution of AI from Tools to Creative Partners

The overwhelming prevalence of request dialogue acts (95% of all interactions) during the narrative writing study suggests that participants primarily approach the AI as a tool to accomplish specific tasks rather than as a collaborator or conversational partner. If the goal is to create AI agents that function as true co-creative partners, we must ask: *What capabilities lead to AI systems that function more as collaborative partners with their own creative contributions and perspectives?*

Future research should investigate how to design AI systems that can function more as true creative partners rather than tools. This could involve developing AI capabilities that go beyond

responding to requests to include proactive contributions, negotiation of creative direction, and even advocacy for particular aesthetic choices. Such systems might bridge the asymmetry observed in current human-AI creative collaborations, potentially enabling new forms of co-creativity that more closely resemble balanced human-human partnerships.

7.3.3 The Tradeoff Between Individual Linguistic Enhancement and Collective Homogenization

The tension between individual linguistic enhancement and collective homogenization represents a fundamental challenge with direct implications for education: While AI assistance can elevate the surface features of student writing, the tendency towards greater semantic similarity raises concerns about the development of unique authorial voice and diverse perspectives. This poses significant challenges for assessment: *How can educators differentiate between genuine improvement in writing skill and effective leveraging of AI tools? Does a higher score on lexical sophistication reflect deeper understanding or simply access to a better tool?* Traditional assessment rubrics focusing solely on the final product may become inadequate.

Future research and pedagogical development must address this tradeoff. This includes exploring assessment methods that focus on the writing process: evaluating students' critical engagement with AI suggestions, their ability to integrate AI contributions coherently, and their judgment in maintaining their own voice and goals.

Furthermore, investigating how different AI architectures or prompting strategies might mitigate homogenization while still providing beneficial support is an important technical and human-computer interaction design challenge. A central question for the design of these future systems is: *How might we design systems that enhance individual creative capabilities without constraining diversity across creators?* Answering this requires exploring innovations in interaction techniques, model transparency, and user control mechanisms that empower creators while leveraging AI strengths.

7.3.4 Long-term Implications for Human Creative Development

The patterns of delegation observed in the narrative writing study raise important concerns about long-term effect of AI integration into creative practices and educational contexts. The substantial proportion of participants (39.1%) who delegated most or all creative responsibility to the AI suggests a potential pathway towards over-reliance that could have long-term implications for human creative skill development and expression, including the development of foundational writing skills such as brainstorming, drafting, revision, and cultivating an authorial voice. As AI tools become increasingly integrated into creative practices and in classrooms, we must ask: *Could over-reliance on AI assistance potentially atrophy certain creative capacities or foundational skills in humans? How does long-term AI assistance affect the development of independent human creative skills? How do patterns of creative agency shift as users become more familiar with AI capabilities? Do users who initially delegate extensively to AI eventually develop more balanced collaboration strategies? and What teaching strategies can effectively leverage AI tools to enhance learning without replacing the essential practice needed for skill acquisition?*

Longitudinal studies tracking how patterns of human-AI creative collaboration evolve over time would provide crucial insights into the long-term developmental implications of AI-assisted creativity.

7.4 Concluding Remarks

This dissertation demonstrates that human-AI co-creativity is not a uniform phenomenon but rather a diverse landscape of engagement patterns, observed across both computational music remixing and narrative writing, each with distinct implications for creative processes and products. The integration of AI into creative practices is not simply adding new capabilities but transforming the creative process itself. The patterns observed—from the polarization in agency distribution to linguistic enhancement coupled with semantic homogenization—suggest that we are witnessing the emergence of new creative paradigms that challenge traditional notions of authorship, originality, and creative identity.

Additionally, the patterns of agency negotiation observed and user expectations of type of support suggest that there is no single answer to how humans collaborate with AI; different individuals conceptualize and engage with AI-assisted creativity in fundamentally different ways. Understanding these differences, and designing systems that respect this diversity, will be essential for realizing the transformative potential of human-AI co-creativity while preserving the richness and diversity of human creative expression.

Ultimately, this research suggests that the most promising path forward lies not in positioning AI as either a replacement for or as a simple enhancement of human creativity, but rather as a complex collaborator whose integration requires careful consideration of interaction dynamics, user agency, and potential trade-offs. By recognizing and seeking to understand both the potential benefits and the limitations of AI assistance, we can work toward co-creative systems that genuinely support human creativity in all its diversity rather than inadvertently constraining it.

APPENDIX A

DIALOGUE ACT TAXONOMY FOR HUMAN-HUMAN CO-CREATIVITY

Table A-1. Taxonomy of co-creative dialogue act labels. Tags that occurred with less than 5% probably in all hidden states and that do not appear in the examples presented in this paper are not included in this table.

Dialogue act label	Relative Frequency in Corpus	Description	Examples
Statement (Stmnt)	17.14%	Utterance of info or explanation, or something that exists in the coding workspace	<i>well we also have to make a loop</i>
Social (Soc)	14.11%	A general salutation, off-task comment, or display of remorse. Plays some social function not captured in the other tags	<i>how are you?</i>
Proposal (P)	12.32%	An assertion of creativity, related to code or music, for the partner to consider	<i>we should do some beats in the background</i>
Directive (Dir)	11.55%	An utterance used to set task responsibilities for each or a single partner	<i>We should focus on the custom function first</i>
Confusion (Con)	10.41%	Seeking help, expressing confusion, lack of knowledge, or uncertainty	<i>What are those variables for?</i>
Acknowledgement (Ack)	6.35%	Accepting the content of the previous utterance or series of utterances	<i>yeah</i>
Passing Responsibility (PR)	6.17%	Passing creative or technical choice to partner	<i>Do you know what sounds you would like to use?</i>
Proposal Acceptance (ProposalAccept)	5.67%	Accepting a partner's addition or assertion to the co-creative mental model shared by both partners	<i>yeah jazzand dubstep sounds fine</i>
Positive Feedback (PosFdbk)	5.29%	Positive response relating to something the partner accomplished within the scope of the task	<i>I liked the piano thing you did</i>
Directive Acceptance (DirAccept)	3.97%	Response to a partner accepting the dictation of flow or direction of project	<i>ok i will figure out a make-beat</i>
Non-positive Feedback (NPosFdbk)	2.29%	Non-positive response relating to something incorrectly done by the partner within the scope of the task	<i>it doesnt sound as good as i thought it would</i>

APPENDIX B HIDDEN MARKOV MODEL DIAGRAM

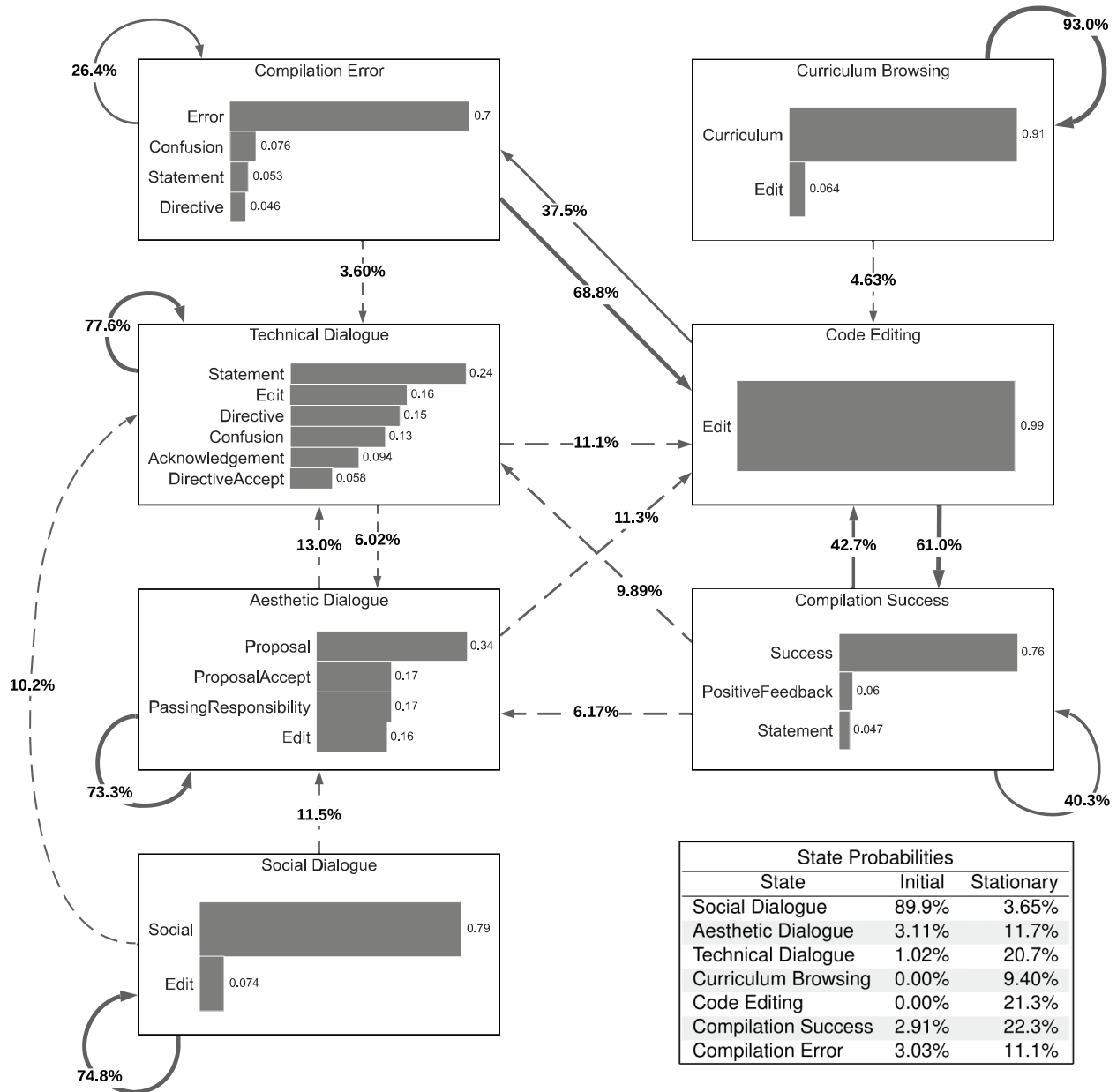


Figure B-1. The learned HMM with $n = 7$ hidden states. Rectangles represent hidden states, and probability distributions within each rectangle indicate emission probabilities. Arrows among hidden states represent transition probabilities. The lower right table displays initial probabilities (probability of a dialogue session beginning in that state) and stationary distribution (proportion of interactions that occur in each state over the long run).

APPENDIX C

CAI DESIGN COMPONENTS

Table C-1. Design elements and capabilities in the third CAI prototype, organized by primary system modules. Changes across versions are noted.

Module	Component	Description
Interface	Chat Window	Dialogue selection and view of messages between user and CAI
	Menu Navigation ^b	Structured topic-specific menus (Music, Help, Controls) enabling direct access to CAI capabilities
	Message Notification ^a	The CAI/Curriculum toggle button flashes when CAI utters a message
Analysis	Complexity Calculator	Evaluates student projects based on demonstrated knowledge of EarSketch curriculum concepts
	Sound Profile	Converts track listing into a representation of sounds and effects by section and measure
Code Suggestion	Weighted Decision-Making ^b	Adaptive suggestion system using user actions and historical effectiveness to select and prioritize recommendations
Dialogue	Dialogue Tree	Menu-based interactions to navigate a tree representing creative phases of student-student pairs
	Project Modeling ^a	Allows users to set project specifications to tune CAI suggestions
	Sound Recommendation ^b	Enhanced system incorporating key signatures and beat complexity for more musically compatible suggestions
	Curriculum Links ^a	Adds interactive links to code suggestions that navigate to relevant Curriculum pages
Error Handling	Code Analysis ^b	Enhanced parsing system for context-aware error interpretation
	Solutions ^b	Expanded debugging suggestions integrating console errors and solution generation

^a Second prototype changes

^b Third prototype changes

APPENDIX D WARMUP ACTIVITIES

D.1 AI Assisted Warmup

Brainstorm Directions

 *For this brainstorm activity, you have the option to interact with an AI agent via the chat window on the right. Note: Please do not use any other tools or resources to aid your writing.*

In the upcoming writing activity, you will be given the beginning of a story involving a hiker's encounter during a late-night walk and asked to continue it. **Before you begin writing your story, please take 3—4 minutes to think about potential characters. Try to come up with 4-5 different characters. Jot down your ideas in the space provided below.** Don't worry about spending too much time on each character; just focus on generating a variety of possibilities. When you write your story, you are free to include, modify, further develop, or disregard any of your brainstormed characters in your story if you wish. Use your creativity and imagination. There are no wrong answers.

After 3 minutes, you may click the "Continue" button below to move on to the story writing activity. After 4 minutes, you will not be able to make changes or chat until you move on to the next activity.

Some aspects you might want to consider for each character:

- Name
- Physical appearance
- Major life events
- Likes and dislikes
- Strengths and flaws

Figure D-1. Warmup Activity: AI Assisted Condition

D.2 Individual Warmup

Brainstorm Directions

 *This brainstorm activity is designed to reflect your personal work. Please do not use any external resources, AI tools, or assistance while completing it.*

In the upcoming writing activity, you will be given the beginning of a story involving a librarian finding a book on their desk one morning and asked to continue it. **Before you begin writing your story, please take 3—4 minutes to think about potential characters. Try to come up with 4-5 different characters. Jot down your ideas in the space provided below.** Don't worry about spending too much time on each character; just focus on generating a variety of possibilities. When you write your story, you are free to include, modify, further develop, or disregard any of your brainstormed characters in your story if you wish. Use your creativity and imagination. There are no wrong answers.

After 3 minutes, you may click the "Continue" button below to move on to the story writing activity. After 4 minutes, you will not be able to make changes or chat until you move on to the next activity.

Some aspects you might want to consider for each character:

- Name
- Physical appearance
- Major life events
- Likes and dislikes
- Strengths and flaws

Figure D-2. Warmup Activity: Individual Condition

APPENDIX E WRITING PROMPTS

E.1 AI Assisted Condition Writing Prompt

Writing Activity Directions

 *For this writing activity, you have the option to interact with an AI agent via the chat window on the right as you complete this activity. Note: Please do not use any other tools or resources to aid your writing.*

Continue the short story, adding approximately 300-500 words. This is the beginning:

Waking up in their tent, a hiker felt restless and found it impossible to drift back to sleep. Seeking some peace, they decided to explore, led by the faint light of their flashlight and the soothing sound of leaves under their feet through the trees. Before long, the trees thinned, and the moonlight revealed a serene clearing with a lake at its center. As they approached the water's edge, they realized the water did not reflect the stars in the sky or the world around it.

After 12 minutes, you may click the *Continue* button below to move on to the post-activity questionnaire. After 16 minutes, you will not be able to make changes in the text editor or chat.

Figure E-1. Writing Activity: AI Assisted Condition

E.2 Individual Condition Writing Prompt

Writing Activity Directions

 *This writing activity is designed to reflect your personal work. Please do not use any external resources, AI tools, or assistance while completing it.*

Continue the short story, adding approximately 300-500 words. This is the beginning:

Arriving at work was the most peaceful part of a librarian's day. The smell of freshly brewed coffee mixed with the musty scent of the books first thing in the morning was a simple pleasure, brightening the start of the day. One morning, as a librarian sat down at their desk, coffee mug in hand, they noticed an unfamiliar book lying there. Upon closer inspection, they discovered their own name and date of birth engraved on the front cover with no other identifying markings or text.

After 12 minutes, you may click the *Continue* button below to move on to the post-activity questionnaire. After 16 minutes, you will not be able to make changes in the text editor.

Figure E-2. Writing Activity: Individual Condition

APPENDIX F

POST-ACTIVITY QUESTIONNAIRE

Table F-1. Post-Activity Evaluation. *Likert Statements*: Completed after each activity. Each statement will be rated with one of the following choices: Strongly Disagree, Disagree, Neither Agree nor Disagree, Agree, Strongly Agree. *Comparative Questions*: Completed after the second post-activity Likert questionnaire. Asks participants to choose between individual and AI conditions. Adapted from [82].

Metric	Likert Statement	Comparative Question
Fun	Writing the short story was fun	What was the most fun way to write a short story?
Enjoyment	I would like to write short stories in this way in the future	With what method would you prefer to write short stories in the future?
Expressiveness (Creativity)	I was able to be creative	With what method were you able to be the most creative?
Expressiveness (Self-Expression)	I was able to express my own thoughts well	With what method were you able to best express yourself?
Outcome Satisfaction	I am happy with the short story I wrote	With which method did you write the short story you were most satisfied with?
Ease of writing (Starting)	It was easy to start writing	With which method was it easiest to start writing a short story?
Ease of writing (Finishing)	Writing was easy	With what method was it easiest to finish a short story?
Ownership	The finished short story is mine	With which method did you write the short story you felt was most yours?

APPENDIX G CREATIVE SELF-EFFICACY AND CREATIVE PERSONAL IDENTITY QUESTIONNAIRE

Table G-1. Questions about creative self-efficacy (CSE) and creative personal identity (CPI)[86].
Participants rated each statement from “Definitely Not” to “Definitely Yes.”

Statement	Type
I think I am a creative person	CPI
My creativity is important for who I am	CPI
I know I can efficiently solve even complicated problems	CSE
I trust my creative abilities	CSE
My imagination and ingenuity distinguishes me from my friends	CSE
Many times I have proved that I can cope with difficult situations	CSE
Being a creative person is important to me	CPI
I am sure I can deal with problems requiring creative thinking	CSE
I am good at proposing original solutions to problems	CSE
Creativity is an important part of myself	CPI
Ingenuity is a characteristic that is important to me	CPI

APPENDIX H

WRITING ATTITUDES AND WRITING SELF-EFFICACY QUESTIONNAIRE

Table H-1. Questions assessing writing attitudes (WA) and writing self-efficacy (WSE) [63, 159].
Participants rated each statement from “Definitely Not” to “Definitely Yes.”

Statement	Type
I usually enjoy writing	WA
The process of writing is satisfying for me	WA
I think that writing is interesting	WA
I try to avoid writing as much as possible	WA
I like to write at home	WA
Writing is a good way to spend my time	WA
I can think of many ideas for my writing	WSE
I can put my ideas into writing	WSE
I can think of many words to describe my ideas	WSE
I can think of a lot of original ideas	WSE
I know exactly where to place my ideas in my writing	WSE

APPENDIX I

LANGUAGE BACKGROUND QUESTIONNAIRE

Table I-1. Language Background Questionnaire derived from The Language and Social Background Questionnaire [6].

Question	Response Options
How many languages do you speak?	1; 2; 3 or more
Is English your first language? ¹	Yes; No
What is your first language? ²	Open text response
Please indicate which language(s) you most frequently heard or used in the following life stages, both inside and outside home: ¹	
Infancy; Preschool Age; Primary School age; High School Age	All English; Mostly English; Half English half other language; Mostly other language; Only other language
In the past 12 months, of the time you spend engaged in each of the following activities, how much of that time is carried out in English? ¹	
Speaking; Listening; Reading; Writing	None; Little; Some; Most; All

¹ Only shown if participant selected the 2 and 3 or more responses for the "How many languages do you speak?" question.

² Only shown if participant indicates English is not their first language.

APPENDIX J
HIERARCHICAL CODING SCHEMA FOR REQUESTS IN AI-ASSISTED WRITING

Table J-1. Primary Dialogue Acts

Label	Definition	Indicators	Examples
Request	User asks the AI to perform a task, provide information, or give feedback.	Imperative verbs, questions, direct or indirect requests	"Write a paragraph about dolphins." "What's the difference between irony and sarcasm?" "Can you review this paragraph?"
Clarification/ Elaboration	User provides additional information or explanation (without making a request).	User explaining their own meaning, adding context to previous statements, correcting misunderstandings	"What I meant was the protagonist should be more reserved." "The story is set in the 1920s." "Actually, I was thinking of a first-person perspective."
Assessment	User provides evaluation or judgment about AI-generated or their own content.	Expressions of approval/disapproval, evaluative statements	"I like the way you described the setting." "This dialogue doesn't sound natural." "The pacing feels too slow here."
Other	Any communication that doesn't fall in the above categories.		"Thank you." "Let's continue tomorrow." "I'm taking a break now."

Table J-2. Request Subcode Taxonomy Table

Category	Label	Definition	Examples
Request Type	Content Generation	User asks the AI to perform a specific task or create content.	"Write the next scene." "Create a character description for the villain."
	Information Seeking	User asks the AI for facts, explanations, or knowledge.	"What's the average rainfall in Seattle?" "How does foreshadowing work in mystery novels?"
Request Directness	Evaluation Request	User asks the AI to evaluate, assess, or provide perspective on content.	"Does this paragraph flow well?" "Is this dialogue realistic for a teenager?"
	Direct	Clearly stated with direct language where the request illocutionary act is directly derivable from the utterance.	"Write the next paragraph." "Explain how metaphors work."
	Indirect	The request is embedded or implied rather than directly stated, requiring pragmatic inference to understand the intended request.	"This part is confusing." "It would be nice if the story had more tension."
Request Constraint	Bounded	User provides specific constraints or parameters that limit the scope of the AI's response.	
	Unbounded	User provides minimal constraints, allowing broad latitude in the AI's response. <i>Requests where all constraints originate from the instructions and/or prompt are considered Open.</i>	
Request Focus	Character	Creating or refining character traits, backgrounds, motivations, or descriptions.	"Create a character profile for the protagonist." "How can I make this villain more sympathetic?"
	Plot	Developing events, actions, conflicts, or narrative progression.	"What should happen next in the story?" "Write a scene where the protagonist confronts the antagonist."
	Setting	Establishing or enhancing environment, time period, or atmosphere.	"Describe the castle where the story takes place." "How can I make my futuristic city feel more lived-in?"
	Thematic	Exploring underlying messages, motifs, or meanings.	"How can I strengthen the theme of betrayal in this story?" "What symbolic elements could I add to reinforce the idea of rebirth?"
	Stylistic	Improving tone, voice, word choice, or literary techniques.	"Make this paragraph more suspenseful." "Rewrite this in a more formal tone."
	Structural	Organizing content, formatting, or presentation of text elements.	"Create an outline for this essay." "Help me with the formatting of this dialogue."
	Informational	Obtaining factual information, explanations, research, or guidance on writing approach/methodology.	"What were common jobs in Victorian England?" "How do I write a compelling opening chapter?"
Request Iteration	New	The request introduces a completely new task or question not previously addressed.	"Write a paragraph about the main character." (When no previous request about the character exists)
	Repeated	The request substantially repeats a previous request, possibly with different wording.	"I asked you to write a scene where the characters meet for the first time." "Let me try again: Can you describe the castle?"
	Refined	The request builds upon a previous request by adding specifications or changing parameters.	"In the scene you wrote, can you add more tension?" "That character description is good, but can you make them more morally ambiguous?"
Content Operation (Only applies to Content Generation Requests)	Generate new content	Request to contribute completely new elements to the story.	"Write a paragraph describing the main character." "Create a scene where two characters meet for the first time."
	Extend user content	Extending or adding on to a previous contribution made by the user.	"I've written the introduction. Can you continue the story from there?"
	Extend AI content	Extending or adding on to a previous contribution made by the agent.	"Continue the scene you just wrote." "Add more details to the setting you described."
	Refine user content	Request asks AI to improve or modify content that the user has created.	"Here's my paragraph. Can you make it more concise?" "Edit my dialogue to sound more natural."
	Refine AI content	Request asks AI to improve or modify content that the AI previously created.	"The paragraph you wrote is too formal. Make it more casual."

APPENDIX K

QUESTIONNAIRE ATTENTION CHECK QUESTIONS

K.1 Attention Check Measures

To ensure data quality and participant engagement throughout the study, I incorporated a series of six attention check questions (AQ1-AQ6) strategically placed within the questionnaires. These attention checks were designed to verify that participants were reading and responding thoughtfully to questionnaire items rather than selecting responses at random. There were several types of attention check questions, detailed in Table K-1, to avoid predictability and strengthen validity [42].

Table K-1. Description of Attention Check Questions

Label	Type	Description
AQ1	Directive check	Instructed participants to select “Somewhat Not” for a specific question. Located in the attitudes questionnaire.
AQ2	Plausibility check	Asked participants to indicate agreement with the impossible statement: “I have never written anything in my entire life.” The correct response was disagreement.
AQ3	Recognition check	Asked about usage of the fictitious AI model “QuantumGPT.” The correct response was “I do not know this model” (full credit) or “I have heard about it, but have not used it” (partial credit).
AQ4	Factual verification	Asked “With which method did you have a chat box available?” The correct response was “AI Assisted,” as chat functionality was only available in the AI-assisted condition.
AQ5	Logical consistency	Presented the statement “Writing was difficult” and expected a response logically opposite to the participant’s earlier answer to “Writing was easy.” For example, if a participant selected “Strongly Agree” that writing was easy, they should select “Strongly Disagree” that writing was difficult.
AQ6	Directive check	Instructed participants to select “Strongly Disagree” for a specific question. Located in the library questionnaire.

The attention checks covered various question formats and were distributed across different sections of the study to maintain consistent attention monitoring throughout the participant experience.

K.2 Scoring Methodology

I implemented a nuanced scoring system for attention checks (1.0 points for correct answers, 0.5 for acceptable but imperfect responses, and 0.0 for incorrect answers). This approach acknowledged that some items might have reasonable alternative interpretations while still demonstrating attention. For example, partial credit was awarded for claiming recognition of, but not using, the fictitious AI model in AQ3, or providing nearly opposite responses in the logical consistency check in AQ5.

K.3 Attention Check Results

The attention check analysis revealed generally high compliance, with a mean score of 5.46 out of 6 possible points (SD = 0.903) across the 23 participants. The majority of participants (69.6%, $n = 16$) successfully passed all attention checks with full credit. Five participants (21.7%) received partial credit (0.5 points) on a single attention check while still answering all others correctly, and two participants (8.7%) earned 5 out of 6 total points. The remaining three participants (13.0%) scored below 4 points, indicating multiple attention check failures.

Question-level performance (Table K-2) varied slightly, with pass rates ranging from 73.9% (AQ5) to 95.7% (AQ1 and AQ4). The logical consistency check (AQ5) proved most challenging, though still with high overall compliance. The fictitious AI model check (AQ3) showed the highest partial credit rate (21.7%), supporting the decision to implement nuanced scoring.

Based on these results, I excluded the three participants (IDs 31, 64, and 81) who scored below 4 points on the attention quality ratings, as their responses indicated potential inattentive or careless responding. Detailed examination of their response patterns confirmed systematic issues with attention check compliance:

Table K-2. Attention Check Response Rates (%) by Question

	AQ1	AQ2	AQ3	AQ4	AQ5	AQ6
Pass Rate (%)	95.65	91.3	78.26	95.65	73.91	91.3
Partial Credit Rate (%)	0.00	0.0	21.74	0.00	17.39	0.0
Fail Rate (%)	4.35	8.7	0.00	4.35	8.70	8.7

- Participant 31 incorrectly answered AQ2 (claiming to have never written anything), received partial credit on AQ3 (claiming knowledge of the fictitious model), and selected the same response for the contradictory statements (AQ5).
- Participant 64 claimed knowledge of the fictitious model (AQ3), selected “Agree” responses for both contradictory statements (AQ5) and selecting “Somewhat Agree” when instructed to select “Strongly Disagree” on AQ6.
- Participant 81 selected incorrect responses on both of the *Directive check* questions (AQ1 and AQ6), and selected the “Individual” option when asked to select the condition with the chat window (AQ4).

Further examination of these participants’ non-attention check responses revealed additional patterns suggesting careless or inconsistent responding. These patterns included extreme response styles (selecting the same answer repeatedly regardless of question content) and logically inconsistent answer combinations (such as preferring individual writing while simultaneously rating AI-assisted writing as superior across multiple dimensions). These response patterns, combined with the attention check failures, provided strong evidence of insufficient engagement with the survey instrument, further supporting the decision to exclude these participants from analysis.

Combined with the previously excluded participants (ID 47 for using generative AI during the individual condition and ID 112 for not using the AI agent at all during the AI-assisted condition), this exclusion resulted in a final analysis sample of 20 participants. This approach aligns with current methodological recommendations to balance data quality concerns with sample retention considerations[2, 128].

LIST OF REFERENCES

- [1] H. Akaike. 1974. A new look at the statistical model identification. *IEEE Trans. Automat. Control* 19, 6 (1974), 716–723.
- [2] Derek A. Albert and Daniel Smilek. 2023. Comparing attentional disengagement between Prolific and MTurk samples. 13, 1 (2023), 20574. doi:[10.1038/s41598-023-46048-5](https://doi.org/10.1038/s41598-023-46048-5)
- [3] Safinah Ali, Hae Won Park, and Cynthia Breazeal. 2020. Can Children Emulate a Robotic Non-Player Character’s Figural Creativity?. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (Virtual Event Canada, 2020-11-02). ACM, 499–509. doi:[10.1145/3410404.3414251](https://doi.org/10.1145/3410404.3414251)
- [4] Alberto Alvarez, Eric Grevillius, Elin Olsson, and Jose Font. 2021. Questgram [Qg]: Toward a Mixed-Initiative Quest Generation Tool. In *Proceedings of the 16th International Conference on the Foundations of Digital Games* (New York, NY, USA, 2021-10-21) (FDG ’21). Association for Computing Machinery, 1–10. doi:[10.1145/3472538.3472544](https://doi.org/10.1145/3472538.3472544)
- [5] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI ’19)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:[10.1145/3290605.3300233](https://doi.org/10.1145/3290605.3300233)
- [6] John A. E. Anderson, Lorinda Mak, Aram Keyvani Chahi, and Ellen Bialystok. 2018. The language and social background questionnaire: Assessing degree of bilingualism in a diverse population. 50, 1 (2018), 250–263. doi:[10.3758/s13428-017-0867-9](https://doi.org/10.3758/s13428-017-0867-9)
- [7] Ivon Arroyo, Naomi Wixon, Danielle Allessio, Beverly Woolf, Kasia Muldner, and Winslow Burleson. 2017. Collaboration Improves Student Interest in Online Tutoring. In *Artificial Intelligence in Education*. Springer International Publishing, 28–39.
- [8] Mitsuru Ikeda Kevin D Ashley and Tak-Wai Chan. 1982. Intelligent tutoring systems. (1982).
- [9] Suyun Sandra Bae, Oh-Hyun Kwon, Senthil Chandrasegaran, and Kwan-Liu Ma. 2020. Spinneret: Aiding Creative Ideation through Non-Obvious Concept Associations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu HI USA, 2020-04-21). ACM, 1–13. doi:[10.1145/3313831.3376746](https://doi.org/10.1145/3313831.3376746)
- [10] Robert F Bales and Fred L Strodbeck. 1951. Phases in group problem-solving. *The Journal of Abnormal and Social Psychology* 46, 4 (1951), 485. doi:[10.1037/h0059886](https://doi.org/10.1037/h0059886)
- [11] Daniel E Berlyne. 1978. Curiosity and learning. *Motivation and emotion* 2, 2 (1978), 97–175. doi:[10.1007/BF00993037](https://doi.org/10.1007/BF00993037)
- [12] Peng Bi. 2020. Revisiting genre effects on linguistic features of L2 writing: A usage-based perspective. 30, 3 (2020), 429–444. doi:[10.1111/ijal.12297](https://doi.org/10.1111/ijal.12297) Publisher: John Wiley & Sons, Ltd.

- [13] Xiaojun Bi, Tom Ouyang, and Shumin Zhai. 2014. Both complete and correct? multi-objective optimization of touchscreen keyboard. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2014) (*CHI '14*). Association for Computing Machinery, 2297–2306.
[doi:10.1145/2556288.2557414](https://doi.org/10.1145/2556288.2557414)
- [14] Kristy Elizabeth Boyer, Eun Young Ha, Michael D Wallis, Robert Phillips, Mladen A Vouk, and James C Lester. 2009. Discovering Tutorial Dialogue Strategies with Hidden Markov Models. In *Artificial Intelligence in Education*. IOS Press, 141–148.
[doi:10.3233/978-1-60750-028-5-141](https://doi.org/10.3233/978-1-60750-028-5-141)
- [15] Kristy Elizabeth Boyer, Robert Phillips, Amy Ingram, Eun Young Ha, Michael D Wallis, Mladen A Vouk, and James C Lester. 2011. Investigating the Relationship Between Dialogue Structure and Tutoring Effectiveness: A Hidden Markov Modeling Approach. *International Journal of Artificial Intelligence in Education* 21, 1 (2011), 65–81.
[doi:10.3233/JAI-2011-018](https://doi.org/10.3233/JAI-2011-018)
- [16] Grant Braught, Tim Wahls, and L Marlin Eby. 2011. The case for pair programming in the computer science classroom. *ACM Transactions on Computing Education (TOCE)* 11, 1 (2011), 1–21. [doi:10.1145/1921607.1921609](https://doi.org/10.1145/1921607.1921609)
- [17] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. In *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (Eds.), Vol. 33. Curran Associates, Inc., 1877–1901. https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfc4967418bfb8ac142f64a-Paper.pdf
- [18] Shannon Campe, Jill Denner, Emily Green, and David Torres. 2020. Pair programming in middle school: variations in interactions and behaviors. *Computer Science Education* 30, 1 (2020), 22–46. [doi:10.1080/08993408.2019.1648119](https://doi.org/10.1080/08993408.2019.1648119)
- [19] Dan Carpenter, Andrew Emerson, Bradford W. Mott, Asmalina Saleh, Krista D. Glazewski, Cindy E. Hmelo-Silver, and James C. Lester. 2020. Detecting Off-Task Behavior from Student Dialogue in Game-Based Collaborative Learning. In *Artificial Intelligence in Education*. Springer International Publishing, 55–66.
- [20] Centers for Disease Control and Prevention. 2023. Disability Impacts All of Us Infographic | CDC. <https://www.cdc.gov/ncbddd/disabilityandhealth/infographic-disability-impacts-all.html>
- [21] Edgar Acosta Chaparro, Aybala Yuksel, Pablo Romero, and Sallyann Bryant. 2005. Factors Affecting the Perceived Effectiveness of Pair Programming in Higher Education.. In *Proceedings of the 17th workshop of the Psychology of Programming Interest Group*. 5–18.

- [22] Mia Xu Chen, Benjamin N. Lee, Gagan Bansal, Yuan Cao, Shuyuan Zhang, Justin Lu, Jackie Tsay, Yinan Wang, Andrew M. Dai, Zhifeng Chen, Timothy Sohn, and Yonghui Wu. 2019. Gmail Smart Compose: Real-Time Assisted Writing. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (New York, NY, USA, 2019) (*KDD '19*). Association for Computing Machinery, 2287–2295. doi:10.1145/3292500.3330723
- [23] Ruijia Cheng, Alison Smith-Renner, Ke Zhang, Joel Tetreault, and Alejandro Jaimes-Larrarte. 2022. Mapping the Design Space of Human-AI Interaction in Text Summarization. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (Seattle, United States, 2022-07), Marine Carpuat, Marie-Catherine de Marneffe, and Ivan Vladimir Meza Ruiz (Eds.). Association for Computational Linguistics, 431–455. doi:10.18653/v1/2022.naacl-main.33
- [24] Peggy Chi, Tao Dong, Christian Frueh, Brian Colonna, Vivek Kwatra, and Irfan Essa. 2022. Synthesis-Assisted Video Prototyping From a Document. In *The 35th Annual ACM Symposium on User Interface Software and Technology* (Bend OR USA, 2022-10-29). ACM, 1–10. doi:10.1145/3526113.3545676
- [25] Edwin Chng, Mohamed Raouf Seyam, William Yao, and Bertrand Schneider. 2020. Using Motion Sensors to Understand Collaborative Interactions in Digital Fabrication Labs. In *Artificial Intelligence in Education*. Springer International Publishing, 118–128.
- [26] Yen-Ting Cho, Yen-Ling Kuo, Yen-Ting Yeh, Yen-Yi Huang, and Po-Lun Huang. 2021. IntuModels: Enabling Interactive Modeling for the Novice through Idea Generation and Selection. In *Creativity and Cognition* (Virtual Event Italy, 2021-06-22). ACM, 1–10. doi:10.1145/3450741.3465241
- [27] John Joon Young Chung, Wooseok Kim, Kang Min Yoo, Hwaran Lee, Eytan Adar, and Minsuk Chang. 2022. TaleBrush: Sketching Stories with Generative Pretrained Language Models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2022-04-29) (*CHI '22*). Association for Computing Machinery, 1–19. doi:10.1145/3491102.3501819
- [28] Elizabeth Clark, Anne Spencer Ross, Chenhao Tan, Yangfeng Ji, and Noah A. Smith. 2018. Creative Writing with a Machine in the Loop: Case Studies on Slogans and Stories. In *23rd International Conference on Intelligent User Interfaces* (Tokyo Japan, 2018-03-05). ACM, 329–340. doi:10.1145/3172944.3172983
- [29] Jacob Cohen. 1988. *Statistical Power Analysis for the Behavioral Sciences* (2 ed.). Lawrence Erlbaum Associates. <https://books.google.com/books?id=FnW8tAEACAAJ>
- [30] Nick Collins. 2008. The Analysis of Generative Music Programs. 13, 3 (2008), 237–248. doi:10.1017/S1355771808000332

- [31] Matthew Conway, Steve Audia, Tommy Burnette, Dennis Cosgrove, and Kevin Christiansen. 2000. Alice: lessons learned from building a 3D system for novices. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 486–493.
- [32] N. Dahlbäck, A. Jönsson, and L. Ahrenberg. 1993. Wizard of Oz studies — why and how. *Know.-Based Syst.* 6, 4 (Dec. 1993), 258–266. doi:[10.1016/0950-7051\(93\)90017-N](https://doi.org/10.1016/0950-7051(93)90017-N) Place: NLD Publisher: Elsevier Science Publishers B. V..
- [33] Hai Dang, Karim Benharraq, Florian Lehmann, and Daniel Buschek. 2022. Beyond Text Generation: Supporting Writers with Continuous Automatic Text Summaries. In *The 35th Annual ACM Symposium on User Interface Software and Technology* (Bend OR USA, 2022-10-29). ACM, 1–13. doi:[10.1145/3526113.3545672](https://doi.org/10.1145/3526113.3545672)
- [34] Neil Davidson and Claire Howell Major. 2014. Boundary crossings: Cooperative learning, collaborative learning, and problem-based learning. *Journal on excellence in college teaching* 25, 3/4 (2014), 7–55.
- [35] Nicholas Davis. 2013. Human-computer co-creativity: Blending human and computational creativity. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, Vol. 9. 9–12. Issue: 6.
- [36] Niraj Ramesh Dayama, Kashyap Todi, Taru Saarelainen, and Antti Oulasvirta. 2020. GRIDS: Interactive Layout Design with Integer Programming. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu HI USA, 2020-04-21). ACM, 1–13. doi:[10.1145/3313831.3376553](https://doi.org/10.1145/3313831.3376553)
- [37] E.L. Deci and R.M. Ryan. 2013. *Intrinsic Motivation and Self-Determination in Human Behavior*. Springer US. <https://books.google.com/books?id=M3CpBgAAQBAJ>
- [38] Jill Denner, Linda Werner, Shannon Campe, and Eloy Ortiz. 2014. Pair Programming: Under What Conditions Is It Advantageous for Middle School Students? *Journal of Research on Technology in Education* 46, 3 (2014), 277–296. doi:[10.1080/15391523.2014.888272](https://doi.org/10.1080/15391523.2014.888272)
- [39] Yong Dich, Joseph Reilly, and Bertrand Schneider. 2018. Using Physiological Synchrony as an Indicator of Collaboration Quality, Task Performance and Learning. In *Artificial Intelligence in Education*. Springer International Publishing, 98–110.
- [40] Pierre Dillenbourg. 1999. *What do you mean by 'collaborative learning'?* Oxford:Elsevier.
- [41] Vladimiras Dolgopolas, Valentina Dagienė, Eglė Jasutė, and Tatjana Jevsikova. 2019. Design science research for computational thinking in constructionist education: a pragmatist perspective. *Problemos* 95 (2019), 144–159.
- [42] Benjamin D. Douglas, Patrick J. Ewell, and Markus Brauer. 2023. Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. 18, 3 (2023), e0279720. doi:[10.1371/journal.pone.0279720](https://doi.org/10.1371/journal.pone.0279720)

- [43] Gregory Dyke, David Adamson, Iris Howley, and Carolyn Penstein Rose. 2013. Enhancing Scientific Reasoning and Discussion with Conversational Agents. *IEEE Trans. Learn. Technol.* 6, 3 (2013), 240–247.
- [44] Sidney D’Mello and Art Graesser. 2012. Dynamics of affective states during complex learning. *Learning and Instruction* 22, 2 (2012), 145–157.
[doi:10.1016/j.learninstruc.2011.10.001](https://doi.org/10.1016/j.learninstruc.2011.10.001)
- [45] Maha Elgarf, Sahba Zojaji, Gabriel Skantze, and Christopher Peters. 2022. CreativeBot: a Creative Storyteller robot to stimulate creativity in children. In *INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION* (Bengaluru India, 2022-11-07). ACM, 540–548. [doi:10.1145/3536221.3556578](https://doi.org/10.1145/3536221.3556578)
- [46] Inan Evin, Perttu Hämäläinen, and Christian Guckelsberger. 2022. Cine-AI: Generating Video Game Cutscenes in the Style of Human Directors. 6 (2022), 1–23. Issue CHI PLAY.
[doi:10.1145/3549486](https://doi.org/10.1145/3549486)
- [47] Oliver Ferschke, Diyi Yang, Gaurav Tomar, and Carolyn Penstein Rosé. 2015. Positive Impact of Collaborative Chat Participation in an edX MOOC. In *Artificial Intelligence in Education*, Cristina Conati, Neil Heffernan, Antonija Mitrovic, and M. Felisa Verdejo (Eds.). Springer International Publishing, 115–124.
[doi:10.1007/978-3-319-19773-9_12](https://doi.org/10.1007/978-3-319-19773-9_12)
- [48] Sally A Fincher and Anthony V Robins. 2019. *The Cambridge handbook of computing education research*. Cambridge University Press.
- [49] Jason Freeman, Brian Magerko, Tom McKlin, Mike Reilly, Justin Permar, Cameron Summers, and Eric Fruchter. 2014. Engaging Underrepresented Groups in High School Introductory Computing through Computational Remixing with EarSketch (*SIGCSE ’14*). Association for Computing Machinery, New York, NY, USA, 85–90.
[doi:10.1145/2538862.2538906](https://doi.org/10.1145/2538862.2538906)
- [50] Jason Freeman, Brian Magerko, and Regis Verdin. 2015. EarSketch: A Web-Based Environment for Teaching Introductory Computer Science Through Music Remixing. In *The 46th ACM Technical Symposium on Computer Science Education* (Kansas City, Missouri, USA) (*SIGCSE ’15*). Association for Computing Machinery, New York, NY, USA, 5.
- [51] Jonas Frich, Lindsay MacDonald Vermeulen, Christian Remy, Michael Mose Biskjaer, and Peter Dalsgaard. 2019. Mapping the Landscape of Creativity Support Tools in HCI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) (*CHI ’19*). Association for Computing Machinery, New York, NY, USA, 1–18. [doi:10.1145/3290605.3300619](https://doi.org/10.1145/3290605.3300619)
- [52] J. Fuegi and J. Francis. 2003. Lovelace & Babbage and the creation of the 1843 ’notes’. 25, 4 (2003), 16–26. [doi:10.1109/MAHC.2003.1253887](https://doi.org/10.1109/MAHC.2003.1253887)

- [53] Kaori Fujinami, Mami Kosaka, and Bipin Indurkha. 2018. Painting an Apple with an Apple: A Tangible Tabletop Interface for Painting with Physical Objects. 2, 4 (2018), 1–22. doi:[10.1145/3287040](https://doi.org/10.1145/3287040)
- [54] Asbjørn Følstad and Cameron Taylor. 2019. Conversational Repair in Chatbots for Customer Service: The Effect of Expressing Uncertainty and Suggesting Alternatives. In *Chatbot Research and Design: Third International Workshop, CONVERSATIONS 2019, Amsterdam, The Netherlands, November 19–20, 2019, Revised Selected Papers*. Springer-Verlag, Berlin, Heidelberg, 201–214. doi:[10.1007/978-3-030-39540-7_14](https://doi.org/10.1007/978-3-030-39540-7_14) event-place: Amsterdam, The Netherlands.
- [55] 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* (Glasgow Scotland Uk, 2019-05-02). ACM, 1–12. doi:[10.1145/3290605.3300526](https://doi.org/10.1145/3290605.3300526)
- [56] Katy Ilonka Gero, Vivian Liu, and Lydia Chilton. 2022. Sparks: Inspiration for Science Writing using Language Models. In *Designing Interactive Systems Conference* (Virtual Event Australia, 2022-06-13). ACM, 1002–1019. doi:[10.1145/3532106.3533533](https://doi.org/10.1145/3532106.3533533)
- [57] Victor Giroto, Erin Walker, and Winslow Burleson. 2019. CrowdMuse: Supporting Crowd Idea Generation through User Modeling and Adaptation. In *Proceedings of the 2019 on Creativity and Cognition* (New York, NY, USA, 2019-06-13) (*C&C '19*). Association for Computing Machinery, 95–106. doi:[10.1145/3325480.3325497](https://doi.org/10.1145/3325480.3325497)
- [58] Vlad-Petre Glăveanu. 2011. How are we creative together? Comparing sociocognitive and sociocultural answers. *Theory & Psychology* 21, 4 (2011), 473–492. doi:[10.1177/0959354310372152](https://doi.org/10.1177/0959354310372152)
- [59] Anuradha A Gokhale. 1995. Collaborative learning enhances critical thinking. 7, 1 (1995), 22–30.
- [60] Bradley A Goodman, Frank N Linton, Robert D Gaimari, Janet M Hitzeman, Helen J Ross, and Guido Zarrella. 2005. Using Dialogue Features to Predict Trouble During Collaborative Learning. *User Modeling and User-Adapted Interaction* 15, 1 (2005), 85–134.
- [61] Jamie Gorson, Nicholas LaGrassa, Cindy Hsinyu Hu, Elise Lee, Ava Marie Robinson, and Eleanor O’Rourke. 2021. An Approach for Detecting Student Perceptions of the Programming Experience from Interaction Log Data. In *Artificial Intelligence in Education*, Ido Roll, Danielle McNamara, Sergey Sosnovsky, Rose Luckin, and Vania Dimitrova (Eds.). Springer International Publishing, Cham, 150–164. doi:[10.1007/978-3-030-78292-4_13](https://doi.org/10.1007/978-3-030-78292-4_13)
- [62] Arthur C. Graesser, Stephen M. Fiore, Samuel Greiff, Jessica Andrews-Todd, Peter W. Foltz, and Friedrich W. Hesse. 2018. Advancing the Science of Collaborative Problem Solving. *Psychological Science in the Public Interest* 19, 2 (2018), 59–92. doi:[10.1177/1529100618808244](https://doi.org/10.1177/1529100618808244)

- [63] Steve Graham, Kausalai Wijekumar, Karen R. Harris, Pui-Wa Lei, Evan Fishman, Amber B. Ray, and Julia Houston. 2019. Writing Skills, Knowledge, Motivation, and Strategic Behavior Predict Students’ Persuasive Writing Performance in the Context of Robust Writing Instruction. 119, 3 (2019), 487–510. doi:[10.1086/701720](https://doi.org/10.1086/701720) Publisher: The University of Chicago Press.
- [64] Shuchi Grover. 2020. *Computer Science in K-12: An Az Handbook on Teaching Programming*. Edfinity.
- [65] Shihui Guo, Yubin Shi, Pintong Xiao, Yinan Fu, Juncong Lin, Wei Zeng, and Tong-Yee Lee. 2023. Creative and Progressive Interior Color Design with Eye-tracked User Preference. 30, 1 (2023), 5:1–5:31. doi:[10.1145/3542922](https://doi.org/10.1145/3542922)
- [66] Mark Guzdial. 1991. Teaching programming with music: An approach to teaching young students about logo. *Logo Foundation* (1991).
- [67] Matthew Guzdial, Nicholas Liao, Jonathan Chen, Shao-Yu Chen, Shukan Shah, Vishwa Shah, Joshua Reno, Gillian Smith, and Mark O. Riedl. 2019. Friend, Collaborator, Student, Manager: How Design of an AI-Driven Game Level Editor Affects Creators. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow Scotland Uk, 2019-05-02). ACM, 1–13. doi:[10.1145/3290605.3300854](https://doi.org/10.1145/3290605.3300854)
- [68] Adam Haar Horowitz, Ishaan Grover, Pedro Reynolds-Cuéllar, Cynthia Breazeal, and Pattie Maes. 2018. Dormio: Interfacing with Dreams. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC Canada, 2018-04-20). ACM, 1–10. doi:[10.1145/3170427.3188403](https://doi.org/10.1145/3170427.3188403)
- [69] Idit Harel and Seymour Papert. 1991. *Constructionism*. Ablex Publishing, Westport, CT, US. Pages: xi, 518.
- [70] Clarice Hilton, Nicola Plant, Carlos González Díaz, Phoenix Perry, Ruth Gibson, Bruno Martelli, Michael Zbyszynski, Rebecca Fiebrink, and Marco Gillies. 2021. InteractML: Making machine learning accessible for creative practitioners working with movement interaction in immersive media. In *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology* (Osaka Japan, 2021-12-08). ACM, 1–10. doi:[10.1145/3489849.3489879](https://doi.org/10.1145/3489849.3489879)
- [71] Md Naimul Hoque, Bhavya Ghai, and Niklas Elmqvist. 2022. DramatVis Personae: Visual Text Analytics for Identifying Social Biases in Creative Writing. In *Designing Interactive Systems Conference* (Virtual Event Australia, 2022-06-13). ACM, 1260–1276. doi:[10.1145/3532106.3533526](https://doi.org/10.1145/3532106.3533526)
- [72] Cynthia Howard, Pamela Jordan, Barbara Di Eugenio, and Sandra Katz. 2017. Shifting the load: A peer dialogue agent that encourages its human collaborator to contribute more to problem solving. *International Journal of Artificial Intelligence in Education* 27, 1 (2017), 101–129.

- [73] Cynthia Howard, Pamela Jordan, Barbara Di Eugenio, and Sandra Katz. 2017. Shifting the Load: a Peer Dialogue Agent that Encourages its Human Collaborator to Contribute More to Problem Solving. *International Journal of Artificial Intelligence in Education* 27, 1 (March 2017), 101–129. doi:[10.1007/s40593-015-0071-y](https://doi.org/10.1007/s40593-015-0071-y)
- [74] Yi-Ching (Janet) Huang, Yu-Ting Cheng, Rung-Huei Liang, Jane Yung-jen Hsu, and Lin-Lin Chen. 2021. Thing Constellation Visualizer: Exploring Emergent Relationships of Everyday Objects. 5 (2021), 1–29. Issue CSCW2. doi:[10.1145/3479866](https://doi.org/10.1145/3479866)
- [75] Alexander Ivanov, David Ledo, Tovi Grossman, George Fitzmaurice, and Fraser Anderson. 2022. MoodCubes: Immersive Spaces for Collecting, Discovering and Envisioning Inspiration Materials. In *Designing Interactive Systems Conference* (New York, NY, USA, 2022-06-13) (*DIS '22*). Association for Computing Machinery, 189–203. doi:[10.1145/3532106.3533565](https://doi.org/10.1145/3532106.3533565)
- [76] Mikhail Jacob and Brian Magerko. 2021. Empirically Evaluating Creative Arc Negotiation for Improvisational Decision-making. In *Creativity and Cognition* (Virtual Event Italy, 2021-06-22). ACM, 1–10. doi:[10.1145/3450741.3465263](https://doi.org/10.1145/3450741.3465263)
- [77] 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* (Hong Kong, China) (*DIS '18*). Association for Computing Machinery, New York, NY, USA, 895–906. doi:[10.1145/3196709.3196735](https://doi.org/10.1145/3196709.3196735)
- [78] Jill V. Jeffery and Kristen Wilcox. 2014. ‘How do I do it if I don’t like writing?’: Adolescents’ stances toward writing across disciplines. 27, 6 (2014), 1095–1117. doi:[10.1007/s11145-013-9493-9](https://doi.org/10.1007/s11145-013-9493-9)
- [79] Youngseung Jeon, Seungwan Jin, Patrick C. Shih, and Kyungsik Han. 2021. FashionQ: An AI-Driven Creativity Support Tool for Facilitating Ideation in Fashion Design. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2021-05-07) (*CHI '21*). Association for Computing Machinery, 1–18. doi:[10.1145/3411764.3445093](https://doi.org/10.1145/3411764.3445093)
- [80] Youwen Kang, Zhida Sun, Sitong Wang, Zeyu Huang, Ziming Wu, and Xiaojuan Ma. 2021. MetaMap: Supporting Visual Metaphor Ideation through Multi-dimensional Example-based Exploration. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2021-05-07) (*CHI '21*). Association for Computing Machinery, 1–15. doi:[10.1145/3411764.3445325](https://doi.org/10.1145/3411764.3445325)
- [81] Anjuli Kannan, Karol Kurach, Sujith Ravi, Tobias Kaufmann, Andrew Tomkins, Balint Miklos, Greg Corrado, Laszlo Lukacs, Marina Ganea, Peter Young, and Vivek Ramavajjala. 2016. Smart Reply: Automated Response Suggestion for Email. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY, USA, 2016) (*KDD '16*). Association for Computing Machinery, 955–964. doi:[10.1145/2939672.2939801](https://doi.org/10.1145/2939672.2939801) event-place: San Francisco, California, USA.

- [82] Anna Kantosalo and Sirpa Riihiäho. 2019. Quantifying co-creative writing experiences. 30, 1 (2019), 23–38. doi:[10.1080/14626268.2019.1575243](https://doi.org/10.1080/14626268.2019.1575243) Publisher: Routledge.
- [83] Anna Kantosalo, Jukka Toivanen, Ping Xiao, and Hannu Toivonen. 2014. From Isolation to Involvement: Adapting Machine Creativity Software to Support Human-Computer Co-Creation. In *The Fifth International Conference on Computational Creativity*, Vol. 2014. 1–7.
- [84] Anna Kantosalo and Hannu Toivonen. 2016. Modes for creative human-computer collaboration: Alternating and task-divided co-creativity. In *Proceedings of the seventh international conference on computational creativity* (2016). 77–84.
- [85] Pegah Karimi, Jeba Rezwana, Safat Siddiqui, Mary Lou Maher, and Nasrin Dehbozorgi. 2020. Creative sketching partner: an analysis of human-AI co-creativity. In *Proceedings of the 25th International Conference on Intelligent User Interfaces* (Cagliari Italy, 2020-03-17). ACM, 221–230. doi:[10.1145/3377325.3377522](https://doi.org/10.1145/3377325.3377522)
- [86] Maciej Karwowski, Izabela Lebeda, and Ewa Wiśniewska. 2018. Measuring creative self-efficacy and creative personal identity. 28, 1 (2018), 45–57. Place: Republic of Korea Publisher: Korean Assn for Thinking Development.
- [87] Gloria Ashiya Katuka, Richard T Bex, Mehmet Celepkolu, Kristy Elizabeth Boyer, Eric Wiebe, Bradford Mott, and James Lester. 2021. My Partner was a Good Partner: Investigating the Relationship between Dialogue Acts and Satisfaction among Middle School Computer Science Learners. In *Proceedings of the 14th International Conference on Computer-Supported Collaborative Learning-CSCL 2021*. International Society of the Learning Sciences.
- [88] Gloria Ashiya Katuka, Alexander R. Webber, Joseph B. Wiggins, Kristy Elizabeth Boyer, Brian Magerko, Tom McKlin, and Jason Freeman. 2022. The Relationship between Co-Creative Dialogue and High School Learners’ Satisfaction with Their Collaborator in Computational Music Remixing. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW1, Article 123 (2022), 24 pages. doi:[10.1145/3512970](https://doi.org/10.1145/3512970)
- [89] Ronald T Kellogg and Merilee Krueger. 1991. The Relative Ease of Writing Narrative Text.
- [90] Kevin Gonyop Kim, Richard Lee Davis, Alessia Eletta Coppi, Alberto Cattaneo, and Pierre Dillenbourg. 2022. Mixplorer: Scaffolding Design Space Exploration through Genetic Recombination of Multiple Peoples’ Designs to Support Novices’ Creativity. In *CHI Conference on Human Factors in Computing Systems* (New Orleans LA USA, 2022-04-29). ACM, 1–13. doi:[10.1145/3491102.3501854](https://doi.org/10.1145/3491102.3501854)
- [91] Yea-Seul Kim, Mira Dontcheva, Eytan Adar, and Jessica Hullman. 2019. Vocal Shortcuts for Creative Experts. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2019-05-02) (*CHI ’19*). Association for Computing Machinery, 1–14. doi:[10.1145/3290605.3300562](https://doi.org/10.1145/3290605.3300562)

- [92] Paivi Kinnunen and Beth Simon. 2010. Experiencing Programming Assignments in CS1: The Emotional Toll. In *Proceedings of the Sixth International Workshop on Computing Education Research (Aarhus, Denmark) (ICER '10)*. Association for Computing Machinery, New York, NY, USA, 77–86. doi:[10.1145/1839594.1839609](https://doi.org/10.1145/1839594.1839609)
- [93] Femke Kirschner, Fred Paas, and Paul A. Kirschner. 2011. Task complexity as a driver for collaborative learning efficiency: The collective working-memory effect. *Applied Cognitive Psychology* 25, 4 (2011), 615–624. doi:[10.1002/acp.1730](https://doi.org/10.1002/acp.1730)
- [94] Paul A Kirschner, John Sweller, Femke Kirschner, and Jimmy Zambrano. 2018. From cognitive load theory to collaborative cognitive load theory. *International Journal of Computer-Supported Collaborative Learning* 13, 2 (2018), 213–233. doi:[10.1007/s11412-018-9277-y](https://doi.org/10.1007/s11412-018-9277-y)
- [95] Maria Knobelsdorf and Ralf Romeike. 2008. Creativity as a Pathway to Computer Science. *SIGCSE Bull.* 40, 3 (2008), 286–290. doi:[10.1145/1597849.1384347](https://doi.org/10.1145/1597849.1384347)
- [96] Hyung-Kwon Ko, Subin An, Gwanmo Park, Seung Kwon Kim, Daesik Kim, Bohyoung Kim, Jaemin Jo, and Jinwook Seo. 2022. We-toon: A Communication Support System between Writers and Artists in Collaborative Webtoon Sketch Revision. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology (New York, NY, USA, 2022-10-28) (UIST '22)*. Association for Computing Machinery, 1–14. doi:[10.1145/3526113.3545612](https://doi.org/10.1145/3526113.3545612)
- [97] Janin Koch, Andrés Lucero, Lena Hegemann, and Antti Oulasvirta. 2019. May AI?: Design Ideation with Cooperative Contextual Bandits. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (Glasgow Scotland Uk, 2019-05-02)*. ACM, 1–12. doi:[10.1145/3290605.3300863](https://doi.org/10.1145/3290605.3300863)
- [98] Janin Koch, Nicolas Taffin, Michel Beaudouin-Lafon, Markku Laine, Andrés Lucero, and Wendy E. Mackay. 2020. ImageSense: An Intelligent Collaborative Ideation Tool to Support Diverse Human-Computer Partnerships. 4 (2020), 1–27. Issue CSCW1. doi:[10.1145/3392850](https://doi.org/10.1145/3392850)
- [99] Janin Koch, Nicolas Taffin, Andrés Lucero, and Wendy E. Mackay. 2020. SemanticCollage: Enriching Digital Mood Board Design with Semantic Labels. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference (Eindhoven Netherlands, 2020-07-03)*. ACM, 407–418. doi:[10.1145/3357236.3395494](https://doi.org/10.1145/3357236.3395494)
- [100] K–12 Computer Science Framework. 2016. K-12 Computer Science Framework. <https://k12cs.org/>
- [101] Essi Lahtinen, Kirsti Ala-Mutka, and Hannu-Matti Järvinen. 2005. A Study of the Difficulties of Novice Programmers. In *Proceedings of the 10th Annual SIGCSE Conference on Innovation and Technology in Computer Science Education (Caparica, Portugal) (ITiCSE '05)*. Association for Computing Machinery, New York, NY, USA, 14–18. doi:[10.1145/1067445.1067453](https://doi.org/10.1145/1067445.1067453)

- [102] Chi Kai Lam. 2023. Technology-enhanced creativity in K-12 music education: A scoping review. *International Journal of Music Education* (2023), 02557614231194073.
- [103] J. Richard Landis and Gary G. Koch. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33, 1 (1977), 159–174.
- [104] John D Lee and Katrina A See. 2004. Trust in Automation: Designing for Appropriate Reliance. *Human Factors* (2004).
- [105] Yuyu Lin, Jiahao Guo, Yang Chen, Cheng Yao, and Fangtian Ying. 2020. It Is Your Turn: Collaborative Ideation With a Co-Creative Robot through Sketch. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2020-04-23) (*CHI '20*). Association for Computing Machinery, 1–14. doi:[10.1145/3313831.3376258](https://doi.org/10.1145/3313831.3376258)
- [106] Yuyu Lin, Jiahao Guo, Yang Chen, Cheng Yao, and Fangtian Ying. 2020. It Is Your Turn: Collaborative Ideation With a Co-Creative Robot through Sketch. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:[10.1145/3313831.3376258](https://doi.org/10.1145/3313831.3376258)
- [107] Fang Liu, Xiaoming Deng, Jiancheng Song, Yu-Kun Lai, Yong-Jin Liu, Hao Wang, Cuixia Ma, Shengfeng Qin, and Hongan Wang. 2022. SketchMaker: Sketch Extraction and Reuse for Interactive Scene Sketch Composition. 12, 3 (2022), 1–26. doi:[10.1145/3543956](https://doi.org/10.1145/3543956)
- [108] Vivian Liu, Han Qiao, and Lydia Chilton. 2022. Opal: Multimodal Image Generation for News Illustration. In *The 35th Annual ACM Symposium on User Interface Software and Technology* (Bend OR USA, 2022-10-29). ACM, 1–17. doi:[10.1145/3526113.3545621](https://doi.org/10.1145/3526113.3545621)
- [109] Xiaofei Lu. 2012. The Relationship of Lexical Richness to the Quality of ESL Learners’ Oral Narratives. 96, 2 (2012), 190–208. doi:[10.1111/j.1540-4781.2011.01232_1.x](https://doi.org/10.1111/j.1540-4781.2011.01232_1.x)
- [110] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) (*CHI '16*). Association for Computing Machinery, New York, NY, USA, 5286–5297. doi:[10.1145/2858036.2858288](https://doi.org/10.1145/2858036.2858288)
- [111] Brian Magerko, Jason Freeman, Tom Mcklin, Mike Reilly, Elise Livingston, Scott Mccoid, and Andrea Crews-Brown. 2016. EarSketch: A STEAM-Based Approach for Underrepresented Populations in High School Computer Science Education. *ACM Transactions on Computing Education (TOCE)* 16, 4, Article 14 (2016), 25 pages.
- [112] Brian Magerko, Jason Freeman, Tom Mcklin, Mike Reilly, Elise Livingston, Scott Mccoid, and Andrea Crews-Brown. 2016. Earsketch: A steam-based approach for underrepresented populations in high school computer science education. *ACM Transactions on Computing Education (TOCE)* 16, 4 (2016), 1–25.

- [113] Neil Maiden, Konstantinos Zachos, Amanda Brown, George Brock, Lars Nyre, Aleksander Nygård Tonheim, Dimitris Apsotolou, and Jeremy Evans. 2018. Making the News: Digital Creativity Support for Journalists. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC Canada, 2018-04-21). ACM, 1–11. doi:[10.1145/3173574.3174049](https://doi.org/10.1145/3173574.3174049)
- [114] Sneha Priscilla Makini, Ini Oguntola, and Deb Roy. 2020. Spelling their pictures: the role of visual scaffolds in an authoring app for young children’s literacy and creativity. In *Proceedings of the Interaction Design and Children Conference* (London United Kingdom, 2020-06-21). ACM, 372–384. doi:[10.1145/3392063.3394392](https://doi.org/10.1145/3392063.3394392)
- [115] Akane Matsumae, Ferdi Trihadi Raharja, Quentin Ehkirch, and Yukari Nagai. 2021. How the Co-Creative Process Affects Concept Formation. *Proceedings of the Design Society 1* (2021), 1775–1786. doi:[10.1017/pds.2021.439](https://doi.org/10.1017/pds.2021.439)
- [116] Charlie McDowell, Linda Werner, Heather Bullock, and Julian Fernald. 2002. The Effects of Pair-Programming on Performance in an Introductory Programming Course. In *Proceedings of the 33rd SIGCSE Technical Symposium on Computer Science Education* (Cincinnati, Kentucky) (*SIGCSE '02*). Association for Computing Machinery, New York, NY, USA, 38–42. doi:[10.1145/563340.563353](https://doi.org/10.1145/563340.563353)
- [117] Bruce M. McLaren and Seiji Isotani. 2011. When Is It Best to Learn with All Worked Examples?. In *Artificial Intelligence in Education*, Gautam Biswas, Susan Bull, Judy Kay, and Antonija Mitrovic (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 222–229. doi:[10.1007/978-3-642-21869-9_30](https://doi.org/10.1007/978-3-642-21869-9_30)
- [118] Danielle S. McNamara, Scott A. Crossley, and Philip M. McCarthy. 2010. Linguistic Features of Writing Quality. 27, 1 (2010), 57–86. doi:[10.1177/0741088309351547](https://doi.org/10.1177/0741088309351547) Publisher: SAGE Publications Inc.
- [119] Danielle S McNamara, Arthur C Graesser, Philip M McCarthy, and Zhiqiang Cai. 2014. *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press.
- [120] Luigi Federico Menabrea and Ada King Countess of Lovelace. 1843. *Sketch of the Analytical Engine Invented by Charles Babbage, Esq.* Richard and John E. Taylor.
- [121] Rohan Mitra and Imran Zualkernan. 2025. Music Generation Using Deep Learning and Generative AI: A Systematic Review. 13 (2025), 18079–18106. doi:[10.1109/ACCESS.2025.3531798](https://doi.org/10.1109/ACCESS.2025.3531798)
- [122] David Moher, Alessandro Liberati, Jennifer Tetzlaff, Douglas G Altman, and PRISMA Group*. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine* 151, 4 (2009), 264–269.
- [123] Roxanne Moore, Michael Helms, and Jason Freeman. 2017. STEAM-Based Interventions in Computer Science: Understanding Feedback Loops in the Classroom. In *2017 ASEE Annual Conference & Exposition*. ASEE Conferences, Columbus, Ohio. doi:[10.18260/1-2--28842](https://doi.org/10.18260/1-2--28842)

- [124] Elizabeth K. Morales-Urrutia, José Miguel Ocaña Ch., Diana Pérez-Marín, and Celeste Pizarro-Romero. 2020. Promoting Learning and Satisfaction of Children When Interacting with an Emotional Companion to Program. In *Artificial Intelligence in Education*. Springer International Publishing, 220–223.
- [125] Laurie Murphy, Sue Fitzgerald, Brian Hanks, and Renée McCauley. 2010. Pair Debugging: A Transactive Discourse Analysis. In *Proceedings of the Sixth International Workshop on Computing Education Research (Aarhus, Denmark) (ICER '10)*. Association for Computing Machinery, New York, NY, USA, 51–58. doi:[10.1145/1839594.1839604](https://doi.org/10.1145/1839594.1839604)
- [126] Timothy Neate, Abi Roper, Stephanie Wilson, Jane Marshall, and Madeline Cruice. 2020. CreaTable Content and Tangible Interaction in Aphasia. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu HI USA, 2020-04-21). ACM, 1–14. doi:[10.1145/3313831.3376490](https://doi.org/10.1145/3313831.3376490)
- [127] Andrew A Neath and Joseph E Cavanaugh. 2012. The Bayesian information criterion: background, derivation, and applications. *Wiley Interdisciplinary Reviews: Computational Statistics* 4, 2 (2012), 199–203.
- [128] Alexander Newman, Yuen Lam Bavik, Matthew Mount, and Bo Shao. 2021. Data Collection via Online Platforms: Challenges and Recommendations for Future Research. 70, 3 (2021), 1380–1402. doi:[10.1111/apps.12302](https://doi.org/10.1111/apps.12302) Publisher: John Wiley & Sons, Ltd.
- [129] 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* (New York, NY, USA, 2018-04-19) (*CHI '18*). Association for Computing Machinery, 1–11. doi:[10.1145/3173574.3173629](https://doi.org/10.1145/3173574.3173629)
- [130] Jianmo Ni, Gustavo Hernandez Abrego, Noah Constant, Ji Ma, Keith Hall, Daniel Cer, and Yinfei Yang. 2022. Sentence-T5: Scalable Sentence Encoders from Pre-trained Text-to-Text Models. In *Findings of the Association for Computational Linguistics: ACL 2022* (Dublin, Ireland, 2022-05), Smaranda Muresan, Preslav Nakov, and Aline Villavicencio (Eds.). Association for Computational Linguistics, 1864–1874. doi:[10.18653/v1/2022.findings-acl.146](https://doi.org/10.18653/v1/2022.findings-acl.146)
- [131] Amy Ogan, Samantha Finkelstein, Erin Walker, Ryan Carlson, and Justine Cassell. 2012. Rudeness and rapport: Insults and learning gains in peer tutoring. In *International Conference on Intelligent Tutoring Systems*. Springer, 11–21.
- [132] 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* (New York, NY, USA, 2018-04-21) (*CHI '18*). Association for Computing Machinery, 1–13. doi:[10.1145/3173574.3174223](https://doi.org/10.1145/3173574.3174223)

- [133] Natalie Parde and Rodney D. Nielsen. 2019. AI Meets Austen: Towards Human-Robot Discussions of Literary Metaphor. In *Artificial Intelligence in Education*. Springer International Publishing, 213–219.
- [134] James L. Peterson. 1980. Computer programs for detecting and correcting spelling errors. 23, 12 (1980), 676–687. doi:[10.1145/359038.359041](https://doi.org/10.1145/359038.359041) Place: New York, NY, USA Publisher: Association for Computing Machinery.
- [135] Sophia Ppali, Vali Laloti, Boyd Branch, Chee Siang Ang, Andrew J. Thomas, Bea S. Wohl, and Alexandra Covaci. 2022. Keep the VRhythm going: A musician-centred study investigating how Virtual Reality can support creative musical practice. In *CHI Conference on Human Factors in Computing Systems* (New Orleans LA USA, 2022-04-29). ACM, 1–19. doi:[10.1145/3491102.3501922](https://doi.org/10.1145/3491102.3501922)
- [136] Lawrence Rabiner and Biinghwang Juang. 1986. An introduction to hidden Markov models. *IEEE ASSP Magazine* 3, 1 (1986), 4–16.
- [137] Iulian Radu, Ethan Tu, and Bertrand Schneider. 2020. Relationships Between Body Postures and Collaborative Learning States in an Augmented Reality Study. In *Artificial Intelligence in Education*. Springer International Publishing, 257–262.
- [138] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *J. Mach. Learn. Res.* 21, 1, Article 140 (jan 2020), 67 pages.
- [139] Naveen Ram, Tanay Gummadi, Rahul Bhethanabotla, Richard J Savery, and Gil Weinberg. 2021. Say What? Collaborative Pop Lyric Generation Using Multitask Transfer Learning. In *Proceedings of the 9th International Conference on Human-Agent Interaction* (Virtual Event Japan, 2021-11-09). ACM, 165–173. doi:[10.1145/3472307.3484175](https://doi.org/10.1145/3472307.3484175)
- [140] Randi Reppen, Nancy Ide, and Keith Suderman. 2005. American National Corpus (ANC) Second Release. doi:[10.35111/251H-G440](https://doi.org/10.35111/251H-G440) Artwork Size: 5662310 KB Pages: 5662310 KB.
- [141] Jenni Reid. 2023. Paul McCartney says A.I. got John Lennon’s voice on ‘last Beatles record’. *CNBC* (June 2023). <https://www.cnbc.com/2023/06/13/paul-mccartney-says-ai-got-john-lennons-voice-on-last-beatles-record.html>
- [142] Fernando J. Rodríguez and Kristy Elizabeth Boyer. 2015. Discovering Individual and Collaborative Problem-Solving Modes with Hidden Markov Models. In *Artificial Intelligence in Education*. Springer International Publishing, 408–418.
- [143] Fernando J Rodríguez, Kimberly Michelle Price, and Kristy Elizabeth Boyer. 2017. Expressing and addressing Uncertainty: A study of collaborative Problem-solving dialogues. In *Proceedings of the 12th International Conference on Computer Supported Collaborative Learning (CSCL)*. Philadelphia, PA: International Society of the Learning Sciences.

- [144] Jeremy Roschelle. 1992. Learning by Collaborating: Convergent Conceptual Change. *Journal of the Learning Sciences* 2, 3 (1992), 235–276.
[doi:10.1207/s15327809jls0203_1](https://doi.org/10.1207/s15327809jls0203_1)
- [145] Jeremy Roschelle and Stephanie D. Teasley. 1995. The Construction of Shared Knowledge in Collaborative Problem Solving. In *Computer Supported Collaborative Learning*, Claire O'Malley (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 69–97.
[doi:10.1007/978-3-642-85098-1_5](https://doi.org/10.1007/978-3-642-85098-1_5)
- [146] Yigal Rosen. 2015. Computer-based Assessment of Collaborative Problem Solving: Exploring the Feasibility of Human-to-Agent Approach. *International Journal of Artificial Intelligence in Education* 25, 3 (2015), 380–406.
- [147] Robert Rosenthal, Harris Cooper, Larry Hedges, and others. 1994. Parametric measures of effect size. 621, 2 (1994), 231–244. Publisher: New York.
- [148] Johnny Saldaña. 2015. *The Coding Manual for Qualitative Researchers* (3rd edition ed.). SAGE Publications Ltd, London.
- [149] Robert-Florian Samoilescu, Mihai Dascalu, Maria-Dorinela Sirbu, Stefan Trausan-Matu, and Scott A. Crossley. 2019. Modeling Collaboration in Online Conversations Using Time Series Analysis and Dialogism. In *Artificial Intelligence in Education*. Springer International Publishing, 458–468.
- [150] Sreecharan Sankaranarayanan, Siddharth Reddy Kandimalla, Sahil Hasan, Haokang An, Christopher Bogart, R. Charles Murray, Michael Hilton, Majd Sakr, and Carolyn Rosé. 2020. Agent-in-the-Loop: Conversational Agent Support in Service of Reflection for Learning During Collaborative Programming. In *Artificial Intelligence in Education*, Ig Ibert Bittencourt, Mutlu Cukurova, Kasia Muldner, Rose Luckin, and Eva Millán (Eds.). Springer International Publishing, Cham, 273–278.
[doi:10.1007/978-3-030-52240-7_50](https://doi.org/10.1007/978-3-030-52240-7_50)
- [151] SAS Institute Inc. 2020. *JMP® 15 Fitting Linear Models*. Cary, NC: SAS Institute Inc.
- [152] SAS Institute Inc. 2021. JMP® Pro 15.
https://www.jmp.com/en_us/software/predictive-analytics-software.html
- [153] R Keith Sawyer. 2012. Explaining creativity: The science of human innovation. (2012). Publisher: Oxford University Press.
- [154] Oliver Schmitt and Daniel Buschek. 2021. CharacterChat: Supporting the Creation of Fictional Characters through Conversation and Progressive Manifestation with a Chatbot. In *Creativity and Cognition* (Virtual Event Italy, 2021-06-22). ACM, 1–10.
[doi:10.1145/3450741.3465253](https://doi.org/10.1145/3450741.3465253)
- [155] Bertrand Schneider and Roy Pea. 2014. Toward collaboration sensing. *International Journal of Computer-Supported Collaborative Learning* 9, 4 (2014), 371–395.

- [156] Hugo Scurto, Bavo Van Kerrebroeck, Baptiste Caramiaux, and Frédéric Bevilacqua. 2021. Designing Deep Reinforcement Learning for Human Parameter Exploration. 28, 1 (2021), 1–35. doi:[10.1145/3414472](https://doi.org/10.1145/3414472)
- [157] Prem Seetharaman, Gautham Mysore, Bryan Pardo, Paris Smaragdis, and Celso Gomes. 2019. VoiceAssist: Guiding Users to High-Quality Voice Recordings. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2019-05-02) (*CHI '19*). Association for Computing Machinery, 1–6. doi:[10.1145/3290605.3300539](https://doi.org/10.1145/3290605.3300539)
- [158] Ticha Sethapakdi and James McCann. 2019. Painting with CATS: Camera-Aided Texture Synthesis. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2019-05-02) (*CHI '19*). Association for Computing Machinery, 1–9. doi:[10.1145/3290605.3300287](https://doi.org/10.1145/3290605.3300287)
- [159] Sergey Shevchuk-Hill, Shana Szczupakiewicz, Bella Kofner, and Kristen Gillespie-Lynch. 2023. Comparing narrative writing of autistic and non-autistic college students. 53, 10 (2023), 3901–3915. doi:[10.1007/s10803-022-05516-z](https://doi.org/10.1007/s10803-022-05516-z) Publisher: Springer.
- [160] Yang Shi, Nan Cao, Xiaojuan Ma, Siji Chen, and Pei Liu. 2020. EmoG: Supporting the Sketching of Emotional Expressions for Storyboarding. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2020-04-23) (*CHI '20*). Association for Computing Machinery, 1–12. doi:[10.1145/3313831.3376520](https://doi.org/10.1145/3313831.3376520)
- [161] Evan Shimizu, Matthew Fisher, Sylvain Paris, James McCann, and Kayvon Fatahalian. 2020. Design Adjectives: A Framework for Interactive Model-Guided Exploration of Parameterized Design Spaces. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (Virtual Event USA, 2020-10-20). ACM, 261–278. doi:[10.1145/3379337.3415866](https://doi.org/10.1145/3379337.3415866)
- [162] Nikhil Singh, Guillermo Bernal, Daria Savchenko, and Elena L. Glassman. 2022. Where to Hide a Stolen Elephant: Leaps in Creative Writing with Multimodal Machine Intelligence. (2022), 3511599. doi:[10.1145/3511599](https://doi.org/10.1145/3511599)
- [163] Jason Smith, Mikhail Jacob, Jason Freeman, Brian Magerko, and Tom Mcklin. 2019. Combining collaborative and content filtering in a recommendation system for a web-based daw. In *Proceedings of the International Web Audio Conference*.
- [164] Jason Smith, Erin JK Truesdell, Jason Freeman, Brian Magerko, Kristy Elizabeth Boyer, and Tom McKlin. 2020. Modeling music and code knowledge to support a co-creative ai agent for education. In *Proceedings of the 20th International Society for Music Information Retrieval Conference*. 134–141.
- [165] Jason Smith, Dillon Weeks, Mikhail Jacob, Jason Freeman, and Brian Magerko. 2019. Towards a hybrid recommendation system for a sound library. In *IUI Workshops*.
- [166] Jason Brent Smith, Ashvala Vinay, and Jason Freeman. 2023. The Impact of Salient Musical Features in a Hybrid Recommendation System for a Sound Library. In *Joint Proceedings of the ACM IUI Workshops*.

- [167] Caitlin Snyder, Nicole M. Hutchins, Gautam Biswas, Mona Emara, Bernard Yett, and Shitanshu Mishra. 2020. Understanding Collaborative Question Posing During Computational Modeling in Science. In *Artificial Intelligence in Education*. Springer International Publishing, 296–300.
- [168] Amy Soller. 2001. Supporting Social Interaction in an Intelligent Collaborative Learning System. *International Journal of Artificial Intelligence in Education* 12 (2001), 40–62.
- [169] Amy Soller and Alan Lesgold. 2007. Modeling the process of collaborative learning. In *The Role of Technology in CSCL: Studies in Technology Enhanced Collaborative Learning*, H. Ulrich Hoppe, Hiroaki Ogata, and Amy Soller (Eds.). Springer US, Boston, MA, 63–86. doi:10.1007/978-0-387-71136-2_5
- [170] Sarah Sterman, Evey Huang, Vivian Liu, and Eric Paulos. 2020. Interacting with Literary Style through Computational Tools. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu HI USA, 2020-04-21). ACM, 1–12. doi:10.1145/3313831.3376730
- [171] Robert J Sternberg. 1999. *Handbook of creativity*. Cambridge University Press.
- [172] Zhida Sun, Sitong Wang, Chengzhong Liu, and Xiaojuan Ma. 2022. Metaphoraction: Support Gesture-based Interaction Design with Metaphorical Meanings. 29, 5 (2022), 1–33. doi:10.1145/3511892
- [173] Merrill Swain. 2000. The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. 97–114 pages.
- [174] Long Tan. 2024. *NeoSCA: A Fork of L2 Syntactic Complexity Analyzer, version 0.1.2*. <https://github.com/tanloong/neosca>
- [175] Cansu Tatar and Deniz Eseryel. 2019. A literature review: Fostering computational thinking through game-based learning in K-12. In *The 42nd Annual Convention of The Association for the Educational Communications and Technology*. 288–297.
- [176] Stephanie D. Teasley. 1997. *Talking About Reasoning: How Important Is the Peer in Peer Collaboration?* Springer Berlin Heidelberg, Berlin, Heidelberg, 361–384. doi:10.1007/978-3-662-03362-3_16
- [177] Stergios Tegos, S. Demetriadis, and T. Tsiatsos. 2013. A Configurable Conversational Agent to Trigger Students’ Productive Dialogue: A Pilot Study in the CALL Domain. *International Journal of Artificial Intelligence in Education* 24 (2013), 62–91.
- [178] Anouk ten Peze, Tanja Janssen, Gert Rijlarsdam, and Daphne van Weijen. 2021. Writing creative and argumentative texts: What’s the difference? : Exploring how task type affects students’ writing behaviour and performance. 21, 1 (2021), 1–38. doi:10.17239/L1ESLL-2021.21.01.11 Section: Articles.

- [179] Derek Thompson. 2022. Breakthroughs of the Year. *The Atlantic* (Dec. 2022). <https://www.theatlantic.com/newsletters/archive/2022/12/technology-medicine-law-ai-10-breakthroughs-2022/672390/> Section: Newsletters.
- [180] Erin Truesdell, Jason Smith, Sarah Mathew, Tom McKlin, Gloria Ashiya Katuka, Amanda E. Griffith, Brian Magerko, Jason Freeman, and Kristy Elizabeth Boyer. 2021. Supporting Computational Music Remixing with a Co-Creative Learning Companion. In *Proceedings of the 12th International Conference on Computational Creativity*. 113–121.
- [181] EJ Truesdell, Jason Brent Smith, Sarah Mathew, Gloria Ashiya Katuka, Amanda Griffith, Tom McKlin, Brian Magerko, Jason Freeman, and Kristy Elizabeth Boyer. 2021. Supporting Computational Music Remixing with a Co-Creative Learning Companion. In *Proceedings of the Twelfth International Conference on Computational Creativity*.
- [182] Jennifer Tsan, Collin F. Lynch, and Kristy Elizabeth Boyer. 2018. “Alright, what do we need?”: A study of young coders’ collaborative dialogue. *International Journal of Child-Computer Interaction* 17 (2018), 61–71. doi:10.1016/j.ijcci.2018.03.001
- [183] Jennifer Tsan, Fernando J. Rodríguez, Kristy Elizabeth Boyer, and Collin Lynch. 2018. ”I Think We Should...”: Analyzing Elementary Students’ Collaborative Processes for Giving and Taking Suggestions. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education* (Baltimore, Maryland, USA) (*SIGCSE ’18*). Association for Computing Machinery, New York, NY, USA, 622–627. doi:10.1145/3159450.3159507
- [184] Franziska Tschan. 1995. Communication Enhances Small Group Performance if it Conforms to Task Requirements: The Concept of Ideal Communication Cycles. *Basic and Applied Social Psychology* 17, 3 (1995), 371–393. doi:10.1207/s15324834basp1703_6
- [185] Marek Urban, Filip Děchtěrenko, Jiří Lukavský, Veronika Hrabalová, Filip Svacha, Cyril Brom, and Kamila Urban. 2024. ChatGPT improves creative problem-solving performance in university students: An experimental study. (2024), 105031. doi:10.1016/j.compedu.2024.105031
- [186] Oleksandra Vereschak, Gilles Bailly, and Baptiste Caramiaux. 2021. How to Evaluate Trust in AI-Assisted Decision Making? A Survey of Empirical Methodologies. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (Oct. 2021), 1–39. doi:10.1145/3476068
- [187] S. A. Viswanathan and K. VanLehn. 2018. Using the Tablet Gestures and Speech of Pairs of Students to Classify Their Collaboration. *IEEE Transactions on Learning Technologies* 11, 2 (2018), 230–242.
- [188] Sandra A. Waddock and Brendan D. Bannister. 1991. Correlates of effectiveness and partner satisfaction in social partnerships. *Journal of Organizational Change Management* 4, 2 (1991), 64–79. doi:10.1108/EUM0000000000001192 Publisher: MCB UP Ltd.

- [189] Tianyi Wang, Ke Huo, Pratik Chawla, Guiming Chen, Siddharth Banerjee, and Karthik Ramani. 2018. Plain2Fun: Augmenting Ordinary Objects with Interactive Functions by Auto-Fabricating Surface Painted Circuits. In *Proceedings of the 2018 Designing Interactive Systems Conference* (Hong Kong China, 2018-06-08). ACM, 1095–1106. doi:[10.1145/3196709.3196791](https://doi.org/10.1145/3196709.3196791)
- [190] Yunlong Wang, Priyadarshini Venkatesh, and Brian Y Lim. 2022. Interpretable Directed Diversity: Leveraging Model Explanations for Iterative Crowd Ideation. In *CHI Conference on Human Factors in Computing Systems* (New Orleans LA USA, 2022-04-29). ACM, 1–28. doi:[10.1145/3491102.3517551](https://doi.org/10.1145/3491102.3517551)
- [191] Blake Williford, Abhay Doke, Michel Pahud, Ken Hinckley, and Tracy Hammond. 2019. DrawMyPhoto: Assisting Novices in Drawing from Photographs. In *Proceedings of the 2019 on Creativity and Cognition* (New York, NY, USA, 2019-06-13) (*C&C '19*). Association for Computing Machinery, 198–209. doi:[10.1145/3325480.3325507](https://doi.org/10.1145/3325480.3325507)
- [192] Blake Williford, Matthew Runyon, and Tracy Hammond. 2020. Recognizing perspective accuracy: an intelligent user interface for assisting novices. In *Proceedings of the 25th International Conference on Intelligent User Interfaces* (Cagliari Italy, 2020-03-17). ACM, 231–242. doi:[10.1145/3377325.3377511](https://doi.org/10.1145/3377325.3377511)
- [193] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. Transformers: State-of-the-Art Natural Language Processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Association for Computational Linguistics, Online, 38–45. doi:[10.18653/v1/2020.emnlp-demos.6](https://doi.org/10.18653/v1/2020.emnlp-demos.6)
- [194] Thomas Wolf, Victor Sanh, Julien Chaumond, and Clement Delangue. 2019. TransferTransfo: A Transfer Learning Approach for Neural Network Based Conversational Agents. *CoRR* abs/1901.08149 (2019). arXiv:1901.08149 <http://arxiv.org/abs/1901.08149>
- [195] Xiaotong (Tone) Xu, Rosaleen Xiong, Boyang Wang, David Min, and Steven P. Dow. 2021. IdeateRelate: An Examples Gallery That Helps Creators Explore Ideas in Relation to Their Own. 5 (2021), 1–18. Issue CSCW2. doi:[10.1145/3479496](https://doi.org/10.1145/3479496)
- [196] Georgios N Yannakakis, Antonios Liapis, and Constantine Alexopoulos. 2014. Mixed-initiative co-creativity. (2014). Publisher: Foundations of Digital Games.
- [197] Junnan Yu and Ricarose Roque. 2019. A review of computational toys and kits for young children. *International Journal of Child-Computer Interaction* 21 (2019), 17–36. doi:[10.1016/j.ijccci.2019.04.001](https://doi.org/10.1016/j.ijccci.2019.04.001)

- [198] Chao Zhang, Cheng Yao, Jiayi Wu, Weijia Lin, Lijuan Liu, Ge Yan, and Fangtian Ying. 2022. StoryDrawer: A Child–AI Collaborative Drawing System to Support Children’s Creative Visual Storytelling. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2022-04-29) (*CHI ’22*). Association for Computing Machinery, 1–15. doi:[10.1145/3491102.3501914](https://doi.org/10.1145/3491102.3501914)
- [199] Jianwei Zhang, Marlene Scardamalia, Mary Lamon, Richard Messina, and Richard Reeve. 2007. Socio-cognitive dynamics of knowledge building in the work of 9- and 10-year-olds. *Educational Technology Research and Development* 55, 2 (2007), 117–145. doi:[10.1007/s11423-006-9019-0](https://doi.org/10.1007/s11423-006-9019-0)
- [200] Jiayi Eris Zhang, Nicole Sultanum, Anastasia Bezerianos, and Fanny Chevalier. 2020. DataQuilt: Extracting Visual Elements from Images to Craft Pictorial Visualizations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu HI USA, 2020-04-21). ACM, 1–13. doi:[10.1145/3313831.3376172](https://doi.org/10.1145/3313831.3376172)
- [201] Mingrui Ray Zhang, He Wen, and Jacob O. Wobbrock. 2019. Type, Then Correct: Intelligent Text Correction Techniques for Mobile Text Entry Using Neural Networks. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA, 2019) (*UIST ’19*). Association for Computing Machinery, 843–855. doi:[10.1145/3332165.3347924](https://doi.org/10.1145/3332165.3347924)
- [202] Rui Zhi, Thomas W. Price, Samiha Marwan, Alexandra Milliken, Tiffany Barnes, and Min Chi. 2019. Exploring the Impact of Worked Examples in a Novice Programming Environment. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education* (Minneapolis, MN, USA) (*SIGCSE ’19*). Association for Computing Machinery, New York, NY, USA, 98–104. doi:[10.1145/3287324.3287385](https://doi.org/10.1145/3287324.3287385)
- [203] Baichang Zhong, Qiyun Wang, and Jie Chen. 2016. The impact of social factors on pair programming in a primary school. *Computers in Human Behavior* 64 (2016), 423–431. doi:[10.1016/j.chb.2016.07.017](https://doi.org/10.1016/j.chb.2016.07.017)

BIOGRAPHICAL SKETCH

Amanda E. Griffith was born in Memphis, Tennessee. She pursued her undergraduate studies in Information Systems at the University of Arkansas' Walton College of Business, where she obtained a Bachelor of Science in Business Administration in May 2017. After graduation, Amanda gained experience working as a data analyst.

In August 2020, she began working on her Ph.D. in the Human-Centered Computing program at the University of Florida. Her doctoral research is conducted under the guidance of Dr. Kristy Elizabeth Boyer within the LearnDialogue Group. In the initial three years, Amanda made contributions to a team dedicated to creating a co-creative AI (CAI) partner designed to aid students in learning coding through music creation using EarSketch. Her early research efforts explored the dynamics of human-human collaboration in creative tasks. Her research focus is on understanding the potential of AI in supporting and augmenting human creativity across various domains. She received her Ph.D. in human-centered computing in May 2025.