Human–Artificial Intelligence Collaborative Creation in Graphic Design: An Experimental Study on Creativity, Efficiency, and User Engagement

Vedant Shashikant Vichare[1], Dr. Sukhada Khandge[2]
[1]*Department of Film & Media, Ajeenkya D Y Patil University, Pune*
[2]*Assistant Professor | ISO Coordinator School of Film & Media (SFM) Ajeenkya D Y Patil University, Pune*

***Abstract*—The rapid proliferation of generative artificial intelligence tools across creative industries has prompted urgent questions about their impact on design quality, workflow efficiency, and audience reception. While industry commentary has been abundant, rigorous experimental evidence comparing Human–AI collaborative design with traditional unaided design particularly within non-Western educational contexts remains scarce. This paper addresses that gap through a controlled quasi-experimental study conducted over six weeks with thirty undergraduate graphic design students at Ajeenkya D Y Patil University, Pune, India.**

Participants were divided into two groups: a Control Group (n = 15) working exclusively with conventional digital design tools, and an Experimental Group (n = 15) integrating AI-assisted platforms (Adobe Firefly, Midjourney v6, and Canva AI) alongside traditional software. Creative quality was assessed through Amabile's (1983) Consensual Assessment Technique by three domain-expert evaluators. Operational efficiency was recorded via structured time-tracking logs. User engagement was measured through a 40-respondent survey evaluating attention capture, emotional appeal, and behavioural intent, supplemented by a delayed recall assessment conducted one-week post-exposure.

Drawing on Sweller's (1988) Cognitive Load Theory, Lubart's (2005) computer-as-creative-partner framework, and Boden's (2004) tripartite creativity taxonomy, this paper proposes the HACE Model a novel conceptual framework that maps the relationships between Human Creativity (H), AI Assistance (A), Cognitive Load (C), and Efficiency (E) in collaborative design workflows. The model predicts that AI assistance reduces extraneous cognitive load, thereby freeing intrinsic cognitive resources for higher-order creative processing a prediction corroborated by all three outcome dimensions of the experimental findings.

Results indicate that the Experimental Group achieved statistically significant superiority across all measured dimensions: higher creative quality scores ($M = 3.87$ vs. 3.21; $p = 0.001$; $d = 1.40$), substantially reduced task completion time (31.4 hrs vs. 47.6 hrs; $p < 0.001$; $d = 3.18$), and stronger user engagement composite scores ($M = 4.02$ vs. 3.46; $p < 0.001$). Delayed recall rates were 67.3% for Experimental Group designs versus 48.1% for Control Group designs. Qualitative thematic analysis revealed four emergent themes: role redefinition from maker to director, confidence through iteration, creative paralysis by choice, and prompt engineering as acquired creative skill.

The paper argues that effective Human–AI collaboration in graphic design constitutes a form of distributed creativity in which the human designer's role shifts from primary generator to creative director a transformation that amplifies rather than diminishes human creative agency. Implications for design pedagogy, professional practice standards, and the ethical governance of AI-assisted creative work are discussed. The study contributes the HACE Model as a replicable theoretical instrument for future research in Human–AI creative collaboration.

***Index Terms*—Human–AI collaboration, graphic design, HACE Model, cognitive load theory, generative AI, creativity assessment, user engagement, design education, experimental study, distributed creativity**

I. INTRODUCTION

1.1 Context and Motivation

The emergence of large-scale generative AI systems including text-to-image models such as Midjourney, DALL·E 3, and Adobe Firefly has created a fundamental inflection point in graphic design practice. These systems have made it possible for designers to generate, iterate, and refine visual concepts at a speed and scale that would have been computationally impossible a decade ago, compressing ideation cycles that once took days into processes measurable in minutes. The practical implications of this compression are significant: if AI tools genuinely reduce the time cost of generative design work without compromising or while actively improving the quality of outputs, then the economic and creative dynamics of design practice will be irreversibly altered.

Yet the discourse surrounding AI in design has been characterized more by assertion than by evidence. Proponents argue that AI democratizes creativity, removes technical barriers, and liberates designers to focus on higher-order conceptual work. Critics contend that AI homogenizes aesthetic output, displaces human creative labor, and produces work that lacks the emotional depth and cultural specificity of human-made design. Both positions are often advanced with minimal empirical grounding. This study was motivated by the conviction that the question of what AI collaboration produces in measurable terms, under controlled conditions is both intellectually serious and practically urgent, and that it deserves the rigorous experimental treatment it has not yet consistently received.

Furthermore, most of the published research on AI and creativity has been conducted in North American or Northern European settings, with participants drawn from technology-oriented institutions. The Indian design education context presents a distinctive case: a rapidly growing professional design sector, a culturally specific visual vocabulary with deep historical roots, and an educational system that is simultaneously modernizing its technical infrastructure and grappling with questions of cultural identity in visual communication. Producing experimental evidence from this context contributes to a more globally representative understanding of how Human–AI creative collaboration functions across different cultural and institutional environments.

1.2 Research Gap

A systematic review of the literature (detailed in Section 2) reveals three specific gaps that this study addresses. First, while numerous studies have evaluated the quality of AI-generated design outputs, very few have compared Human–AI collaborative outputs with human-only outputs under controlled experimental conditions, using validated creativity assessment instruments. Second, the efficiency dimension of Human–AI design collaboration has been documented primarily through industry surveys and retrospective self-report; time-tracking data collected prospectively during a controlled experiment provides substantially stronger evidence. Third, the user engagement dimension how audiences actually respond to AI-assisted versus traditionally produced design has received almost no dedicated empirical attention in peer-reviewed literature, despite being the dimension most directly relevant to design's communicative purpose.

Beyond these empirical gaps, there is also a theoretical gap: no existing framework specifically maps the relationships among human creativity, AI assistance, cognitive load management, and efficiency in the context of graphic design. The HACE Model, proposed in this paper, is offered as a contribution to filling this theoretical gap and as a foundation for cumulative research across different design contexts and populations.

1.3 Research Objectives and Questions

The study was organized around three primary research objectives, each associated with a directional research hypothesis:

RO1: To determine whether Human–AI collaborative graphic design produces work of higher assessed creative quality than traditional unaided design. H1: Experimental Group designs will achieve significantly higher scores than Control Group designs on all four dimensions of the Consensual Assessment Technique.

RO2: To quantify the efficiency differential between Human–AI collaborative and traditional design processes across distinct workflow phases. H2: Experimental Group participants will complete identical design tasks in significantly less total time, with the greatest efficiency gains concentrated in the ideation phase.

RO3: To assess whether Human–AI collaborative designs generate stronger user engagement responses than traditionally produced designs. H3: Experimental Group designs will receive significantly higher ratings on attention capture, emotional appeal, and behavioural intent, and will be recalled more accurately after a one-week delay.

1.4 Significance and Scope

The significance of this research extends across three levels. At the theoretical level, the HACE Model provides a new conceptual instrument for understanding the cognitive and creative dynamics of Human–AI collaboration one grounded in established psychological theory (Cognitive Load Theory) and applicable beyond the specific domain of graphic design. At the empirical level, the study provides controlled experimental evidence on a question of considerable practical importance to design practitioners, educators, and tool developers. At the applied level, the findings generate specific, evidence-based recommendations for design pedagogy, professional practice standards, and the development of AI creative tools.

The scope of the study is deliberately focused: it examines undergraduate-level graphic design students in a single institutional context, working with a defined set of AI tools on four standardized design tasks. The findings should be interpreted and generalized within these boundaries, and the limitations section (6.4) addresses how future research might extend the evidence base to broader populations and contexts.

1.5 Structure of the Paper

Section 2 presents a comprehensive literature review organized around four theoretical pillars: Cognitive Load Theory, creativity in design, Human–AI collaboration frameworks, and user engagement with visual communication. Section 3 introduces the HACE Model and its theoretical derivation. Section 4 describes the research methodology in formal academic detail. Section 5 presents and analyzes the experimental results. Section 6 discusses the findings in theoretical and applied context. Section 7 describes the specific research contributions of this paper. Section 8 presents conclusions and recommendations. References and appendices follow.

II. LITERATURE REVIEW

2.1 Cognitive Load Theory and Creative Performance

Cognitive Load Theory (CLT), originally formulated by Sweller (1988) and elaborated in subsequent decades (Sweller, van Merrienboer, & Paas, 1998; Sweller, 2010), posits that the human working memory system operates under finite capacity constraints. CLT distinguishes three types of cognitive load: intrinsic load, which arises from the inherent complexity of the material being processed; extraneous load, which is imposed by the manner in which information is presented or tasks are structured and which does not contribute to learning or creative production; and germane load, which reflects the cognitive effort invested in schema formation and the integration of new information into long-term memory.

The implications of CLT for creative performance are significant. When extraneous cognitive load is high for example, when a designer must devote substantial mental resources to the mechanical execution of a visual concept before that concept can be evaluated working memory resources available for generative creative thinking (germane load) are correspondingly reduced. This theoretical prediction generates a clear hypothesis about the role of AI in design: by automating or accelerating the production of initial visual forms, AI assistance reduces extraneous load, thereby releasing working memory capacity for the generative, evaluative, and directive cognitive processes that constitute creative design thinking.

Paas and van Merriënboer (1994) operationalized cognitive load measurement through mental effort rating scales, and their approach has been widely validated in learning and performance research. In the context of this study, the efficiency data (time-tracking logs) serve as an indirect proxy for cognitive load allocation: reduced time in the ideation and drafting phases, combined with maintained or improved quality, is consistent with a reduction in extraneous load rather than a reduction in creative investment. The qualitative interview data provides complementary direct evidence by examining participants' subjective experience of cognitive demand in the two experimental conditions.

More recently, Kirsh (2010) extended CLT's application to creative practice through the concept of cognitive offloading the use of external representations and tools to reduce the working memory demands of complex cognitive tasks. Sketch-making, note-taking, and physical modelling are classic examples of cognitive offloading in design practice; AI-generated imagery represents a technologically amplified form of the same cognitive strategy, externalizing the visual form of a concept so that the designer's working memory can engage with it evaluatively rather than generatively. This framing is central to the HACE Model developed in Section 3.

2.2 Creativity in Design: Theoretical Foundations and Measurement

Creativity in design is not a unitary construct. Boden's (2004) influential tripartite taxonomy distinguishes combinational creativity (generating novel combinations of existing concepts), exploratory creativity (expanding the boundaries of an established conceptual space), and transformational creativity (fundamentally restructuring the rules of a conceptual space). For graphic design practice, combinational creativity the ability to draw unexpected and effective connections between visual ideas, cultural references, and communicative purposes is the most operationally relevant form, and it is precisely the domain in which AI systems trained on large visual datasets demonstrate the most powerful capability.

Amabile's (1983) componential model of creativity identifies three necessary conditions for creative output: domain-relevant skills (the technical and conceptual knowledge of the field), creativity-relevant processes (cognitive styles and heuristics conducive to creative thinking), and intrinsic task motivation (engagement driven by interest and satisfaction in the work itself). AI assistance has differential implications for each component: it can substitute for some domain-relevant technical skills (reducing the expertise required for visual production), potentially enhances creativity-relevant processes by expanding the space of options explored, and may affect intrinsic motivation either positively (by removing frustrating technical barriers) or negatively (by reducing the sense of personal ownership over the generative process).

The Consensual Assessment Technique (CAT), Amabile's (1983) methodological contribution for measuring creativity in domain-specific tasks, operationalizes creative quality as the consensus judgment of domain experts a standard that is both practically useful and theoretically defensible on the grounds that what a community of expert practitioners agrees is creative is, definitionally, creative within that domain. The CAT has been validated across a wide range of creative domains and remains the gold standard for creativity measurement in experimental research. Its application in this study is detailed in Section 4.

Csikszentmihalyi's (1996) systems model of creativity locates creative quality not in the individual alone but in the interaction among the individual, the domain (the established knowledge and conventions of the field), and the field (the expert community that evaluates and validates creative work). This model is directly relevant to understanding Human–AI collaboration: AI systems, trained on large repositories of human-created work, function as an externalized domain resource a vast repository of domain knowledge available to the individual designer as an interactive partner. The creative quality of the collaboration thus depends on the quality of the designer's interaction with this externalized domain, not merely on the AI's generative capability in isolation.

Divergent thinking, a foundational construct in creativity research since Guilford (1967), refers to the cognitive capacity to generate multiple distinct responses to an open-ended problem a capacity measured through fluency (number of responses), flexibility (variety of categories represented), and originality (statistical infrequency of responses). AI assistance in design can be understood as an external amplifier of divergent thinking, systematically expanding fluency and flexibility by generating visual options beyond the designer's individual cognitive reach. Whether this amplification also produces genuine originality as assessed by expert consensus is an empirical question that the present study directly addresses.

2.3 Human–AI Collaboration: Frameworks and Empirical Evidence

Lubart's (2005) typology of computer roles in creative work tool, consultant, collaborator, and co-creator provides the foundational framework for understanding the spectrum of Human–AI creative relationships. Contemporary generative AI design tools occupy the collaborator position: they respond to human direction (prompts) but also generate outputs that go beyond what the human explicitly specified, contributing genuinely novel visual ideas that the human must then evaluate, select, and integrate into a coherent design. The boundary between collaborator and co-creator is being actively contested as AI model sophistication increases.

Rezwana and Maher (2023) provide the most empirically detailed recent investigation of Human–AI co-creative systems, examining how the distribution of creative agency between human and AI affects creative outcomes and user experience. Their central finding that fluid, role-alternating collaboration produces higher creative satisfaction than fixed hierarchical collaboration has direct pedagogical implications: effective Human–AI design collaboration is a dynamic interactional skill, not a static application of a tool.

Dove et al. (2017) investigated how UX designers conceptualize AI and found that designers' mental models of AI whether they treated it as a tool, a medium, or a collaborator significantly shaped the quality of their collaborative outcomes. Designers who adopted a collaborator mental model produced more creative and more contextually appropriate outputs than those who treated AI as a tool for executing predetermined ideas. This finding suggests that design education must address not only the technical skills of AI-assisted design but the conceptual frameworks through which designers approach the collaboration.

Kantosalo and Toivonen (2016) developed the concept of alternating co-creativity, proposing that human-computer creative collaboration is most productive when leadership of the generative process shifts fluidly between human and machine depending on the nature of the task at hand. Their framework anticipates the qualitative finding of this study (Theme 1: the director's shift) and provides a theoretical basis for understanding how the most effective EG participants structured their AI interactions.

Yang et al. (2023) tracked UX designers' workflow evolution over one year of AI tool adoption and documented a systematic shift in task composition: AI-assisted designers spent proportionally less time on visual production and more time on research, strategy, and stakeholder communication. This workflow transformation which the present study captures in compressed form through the phase-by-phase time analysis represents a fundamental reorientation of the designer's professional role, with significant implications for how design competency is defined and assessed.

The phenomenon of creative paralysis by choice, identified in the qualitative findings of this study, has theoretical antecedents in Schwartz's (2004) paradox of choice research, which demonstrated that beyond a threshold number of options, choice abundance reduces decision satisfaction and increases decision latency. In the design context, the proliferation of AI-generated alternatives can create a form of aesthetic decision paralysis in which the cognitive demands of evaluating and choosing among options exceed the cognitive savings achieved by not having to generate them manually. Managing this paradox is a key metacognitive skill for effective AI-assisted design.

2.4 User Engagement with Visual Communication

Sundar's (2008) MAIN model identifies four dimensions of media engagement: Modality (richness and appropriateness of sensory presentation), Agency (the degree of user control), Interactivity (responsiveness to user input), and Navigability (ease of movement through content). For static graphic design the primary output type in this study Modality is the dominant engagement dimension. Design tools that help designers optimize visual modality through greater aesthetic coherence, compositional boldness, and emotional resonance should therefore have measurable positive effects on audience engagement outcomes.

Tractinsky, Katz, and Ikar's (2000) landmark study established the aesthetic-usability effect: perceived visual beauty significantly predicts positive user responses across multiple dimensions of experience, including perceived quality, trustworthiness, and behavioural engagement. This effect has been replicated in diverse cultural contexts (Reinecke & Bernstein, 2011) and provides a theoretical basis for predicting that designs of higher assessed aesthetic quality such as those produced by the Experimental Group in this study will also generate stronger user engagement responses.

The relationship between visual distinctiveness and memory retention has been extensively documented in cognitive psychology. The Von Restorff effect (von Restorff, 1933, as cited in Hunt, 1995) established that items that are visually or conceptually distinctive within a set are more likely to be encoded in long-term memory and recalled accurately after a delay. The superior recall rates for Experimental Group designs in this study's delayed recall assessment are consistent with this effect: AI-assisted designs, by exploring a broader visual vocabulary and producing more compositionally distinctive outputs, may create stronger mnemonic traces in audience memory.

Kim et al. (2022) documented the attribution effect the tendency for audiences to rate art lower on measures of emotional authenticity when informed it was produced by AI but found that this effect was substantially attenuated when AI was described as a collaborative tool used by a human designer rather than as the sole creator. This finding is directly relevant to the disclosure question addressed in Section 6.3 and suggests that the framing of Human–AI collaboration in professional communication may significantly moderate audience response.

III. THE HACE MODEL: A CONCEPTUAL FRAMEWORK FOR HUMAN–AI CREATIVE COLLABORATION

3.1 Rationale for a New Framework

Existing frameworks for understanding creativity, Human–AI collaboration, and cognitive performance have been developed independently and have not been systematically integrated for the specific context of AI-assisted graphic design. Cognitive Load Theory addresses the relationship between cognitive resources and task performance but was developed for learning contexts rather than creative production. Lubart's (2005) collaboration typology addresses the structure of Human–AI creative relationships but does not incorporate cognitive load mechanisms. Boden's (2004) creativity taxonomy describes types of creative output but does not address the process dynamics of collaborative generation. A framework that integrates these theoretical resources into a unified model specific to the context of Human–AI design collaboration is therefore both theoretically justified and practically needed.

The HACE Model an acronym for Human Creativity, AI Assistance, Cognitive Load, and Efficiency is proposed as this integrative framework. It is not offered as a fully formalized computational model but as a conceptual map of the key relationships among these four constructs, grounded in existing theory and designed to generate testable predictions for empirical research.

3.2 Components of the HACE Model

Component H: Human Creativity

Human Creativity (H) in the HACE Model refers to the designer's capacity for original creative thinking their ability to generate novel, appropriate, and aesthetically distinctive design solutions. Following Amabile (1983), this capacity is understood as the product of three interacting resources: domain expertise (design knowledge and visual culture literacy), creative process skills (divergent thinking, aesthetic judgment, conceptual synthesis), and intrinsic motivation (engagement with and investment in the creative task). H is not a fixed trait but a dynamic capacity whose expression depends on the conditions of the creative context including, critically, the cognitive load imposed by the production process.

Component A: AI Assistance

AI Assistance (A) in the HACE Model refers to the generative, iterative, and evaluative support provided by AI tools within the design workflow. Following Lubart's (2005) collaborator role, A is characterized by its capacity to produce multiple high-quality visual alternatives in response to human direction, to do so at a speed that exceeds human manual production by orders of magnitude, and to draw on a training corpus of visual knowledge that extends far beyond any individual designer's visual experience. A does not possess creative agency in the full sense it cannot initiate a creative direction, evaluate contextual appropriateness, or adapt to culturally specific communicative purposes without human guidance. Its contribution is generative abundance and production velocity.

Component C: Cognitive Load

Cognitive Load (C) in the HACE Model follows Sweller's (1988, 2010) tripartite distinction between intrinsic, extraneous, and germane load. In a traditional (non-AI-assisted) design workflow, extraneous cognitive load is substantially imposed by the mechanical demands of visual production: translating a mental concept into a digital form requires sustained effort in software operation, manual adjustment, and iterative refinement that consumes working memory capacity. In a Human–AI collaborative workflow, much of this extraneous load is offloaded to the AI system (Kirsh, 2010), freeing working memory resources for the germane cognitive processes concept evaluation, aesthetic judgment, directorial decision-making that are most closely associated with creative quality.

Component E: Efficiency

Efficiency (E) in the HACE Model refers to the ratio of creative output quality to the investment of time and cognitive resources. Efficiency is not synonymous with speed: a design produced very quickly but at low quality has low efficiency in this sense, while a design produced in moderate time at high quality has high efficiency. The HACE Model predicts that AI assistance improves efficiency by simultaneously increasing output quality (through expanded creative exploration) and reducing resource investment (through cognitive load offloading). This dual mechanism distinguishes the HACE model's efficiency prediction from simpler accounts that treat AI merely as a production accelerator.

3.3 The HACE Model: Core Propositions

The HACE Model generates four core theoretical propositions, each of which corresponds to a testable empirical prediction:

3.4 The HACE Model and the Present Study

The three experimental hypotheses of this study (H1–H3, Section 1.3) are direct operationalizations of the HACE Model's predictions. H1 (superior creative quality in the Experimental Group) corresponds to HACE Proposition 1. H2 (greater efficiency with largest gains in ideation) corresponds to HACE Propositions 3 and the phase-specific prediction of Proposition 3. H3 (stronger user engagement for Experimental Group outputs) is predicted by the model as a downstream consequence of H1: if Human–AI collaboration produces designs of higher creative quality, those designs should also generate stronger audience responses, consistent with Tractinsky et al.'s (2000) aesthetic-usability effect.

The qualitative finding of creative paralysis by choice among three EG participants directly instantiates HACE Proposition 4's prediction of quality degradation when A overwhelms H when the abundance of AI-generated options exceeds the designer's capacity for decisive aesthetic judgment. This finding is not a threat to the model but a confirmation of its fourth proposition, which predicts that the H-A relationship is not uniformly positive but depends on the quality of the human designer's directorial engagement with the AI's generative output.

IV. RESEARCH METHODOLOGY

4.1 Research Design

This study employed a quasi-experimental, post-test-only design with pre-experiment equivalence verification, following the methodological framework recommended by Campbell and Stanley (1963) for applied educational research contexts where full randomization of pre-existing groups is logistically constrained. Two parallel groups operated independently over a six-week experimental period: a Control Group (CG) working with conventional digital design tools and an Experimental Group (EG) integrating AI-assisted platforms into their workflow. The experimental manipulation AI tool integration constituted the independent variable. The three dependent variables were: (1) assessed creative quality, (2) task completion efficiency (inverse of total time), and (3) user engagement composite score.

A mixed-methods design was adopted for data collection. Quantitative data (creativity scores, time logs, engagement ratings) were analysed through inferential statistics. Qualitative data (post-experiment semi-structured interviews with all 15 EG participants) were analysed through thematic analysis (Braun & Clarke, 2006) and integrated with quantitative findings through a convergent parallel mixed-methods strategy (Creswell & Plano Clark, 2018).

4.2 Participants

Thirty undergraduate students enrolled in the BA JMP and B.Sc. GAD programs at Ajeenkya D Y Patil University, Pune, were recruited through purposive sampling based on four eligibility criteria: (a) sixth-semester enrolment with a minimum of four completed semesters of formal design training; (b) demonstrated proficiency in Adobe Creative Suite (verified through department records); (c) no prior structured experience with generative AI design tools (verified through self-report screening); and (d) voluntary informed consent.

Participants were stratified on two baseline variables standardized design aptitude score and cumulative design GPA and randomly assigned to conditions within strata. Pre-experiment group comparisons confirmed adequate equivalence: design aptitude (CG: $M = 72.4$, $SD = 8.1$; EG: $M = 73.1$, $SD = 7.9$; $t(28) = 0.33$, $p = .74$) and design GPA (CG: $M = 7.6$, $SD =$

0.8; EG: M = 7.7, SD = 0.7; t(28) = 0.38, p = .61). No statistically significant pre-experiment differences were detected on either measure, supporting the attribution of post-experiment differences to the experimental manipulation.

All participants provided written informed consent prior to the study. The research protocol was reviewed and approved by the institutional academic committee. Participants were informed of their right to withdraw at any time without academic consequence and received a full debrief at the study's conclusion.

4.3 Independent Variable: AI Tool Integration

The experimental condition consisted of structured access to and training in three AI design platforms: Adobe Firefly (for style-consistent image generation natively integrated with Adobe Creative Cloud), Midjourney v6 (for high-resolution concept image generation via text prompt), and Canva AI (for rapid layout suggestion and asset generation). These platforms were selected to represent the range of AI capabilities most relevant to professional graphic design practice, and to ensure that EG participants had access to both specialized creative AI and integrated workflow AI.

AI tool training was delivered through a four-session structured workshop series (total: 8 hours) conducted in the week preceding the six-week experimental period. Workshop content covered: (1) principles of prompt engineering for visual AI systems; (2) workflow integration protocols for combining AI-generated assets with Adobe Creative Suite software; (3) critical evaluation of AI outputs for cultural appropriateness, originality, and brief relevance; and (4) ethical practice guidelines for AI-assisted design. The CG received no AI tool training and used only Adobe Illustrator, Photoshop, and InDesign throughout the study.

4.4 Experimental Tasks

Both groups received identical design briefs for four tasks administered sequentially across the six-week period. Tasks were designed to represent the spectrum of professional graphic design practice and to test the experimental variables across different visual formats and communicative contexts.

Table. Task Specifications

Task	Format	Brief Summary	Primary Test	Duration
T1	Social Media Campaign (5 posts, 1080×1080px)	Brand campaign for fictional sustainable fashion label 'Earth Thread'	Creativity + Efficiency	Weeks 1–2
T2	Cultural Event Poster (A2 print format)	Multicultural arts festival 'Rang Utsav' — bilingual (Hindi/English)	Cultural Specificity + Creativity	Weeks 2–3
T3	Product Packaging (front & back panels)	Artisanal Indian spice brand 'Mitti' — premium retail positioning	Aesthetic Quality + Brief Relevance	Week 4
T4	Digital Infographic	Mental health statistics for Indian youth (18–25), social-first format	Data Visualization + Engagement	Weeks 5–6

Note. All four tasks were evaluated on the same four CAT dimensions. Task duration refers to the allocated experimental period; actual completion time was tracked individually.

4.5 Dependent Variable Measurement

4.5.1 Creative Quality (DV1)

Creative quality was operationalized using an adapted Consensual Assessment Technique (CAT; Amabile, 1983). Three expert evaluators professional graphic designers with a minimum of eight years of industry experience, recruited across advertising, editorial, and packaging design sectors independently rated each design on four dimensions using five-point Likert scales:

- Originality: Degree to which the design presents fresh, unexpected, or unconventional visual solutions (1 = entirely conventional; 5 = highly original)
- Aesthetic Quality: Coherence, balance, typographic quality, and compositional sophistication of the design (1 = poor; 5 = excellent)
- Relevance to Brief: Degree to which the design effectively addresses the specific communicative requirements of the task (1 = does not address brief; 5 = fully and compellingly addresses brief)
- Conceptual Strength: Clarity and force of the underlying creative concept unifying all visual elements (1 = no discernible concept; 5 = strong, clear concept)

All designs were evaluated blind to group membership through removal of identifying information and randomization of evaluation sequence. Inter-rater reliability was assessed using Krippendorff's alpha (α = .81 across all dimensions), exceeding the conventional .80 threshold for acceptable agreement (Krippendorff, 2004). Discrepancies of more than one point between any two evaluators on any dimension were resolved through a structured consensus protocol.

4.5.2 Efficiency (DV2)

Efficiency was measured through mandatory daily time-tracking logs submitted by all participants for the duration of the six-week period. Logs recorded active design working time in five categories: ideation, initial drafting, refinement, peer-feedback integration, and finalization. Participants were instructed to log time immediately following each working session. Log accuracy was verified through weekly faculty review and structured reconciliation interviews for three participants who displayed inter-session log anomalies. Total task completion time per participant was used as the primary efficiency metric (lower = more efficient).

4.5.3 User Engagement (DV3)

User engagement was assessed through a two-stage evaluation protocol administered to a 40-person non-designer respondent panel recruited from the university's broader student population. In Stage 1 (immediate evaluation), respondents were each randomly assigned 12 designs (six CG, six EG, group identity concealed) and rated each on three five-point Likert items: Attention Capture, Emotional Appeal, and Likelihood of Behavioural Response. A composite engagement score was calculated as the mean of the three items (Cronbach's α = .83, indicating acceptable internal consistency).

In Stage 2 (delayed recall, administered one-week post-exposure), respondents were presented with 12 thumbnail images six previously seen (three CG, three EG) and six new foils and asked to identify which images they recognized and to describe one specific visual element they recalled from each recognized design. Recognition accuracy and descriptive specificity were coded by two independent raters (Cohen's κ = .79, indicating adequate inter-rater reliability).

4.6 Variables Summary

Table. Summary of Independent and Dependent Variables

Variable Type	Variable	Operationalization	Measurement Instrument
Independent	AI Tool Integration	Presence/absence of AI design tools in workflow	Group assignment (CG vs. EG)
Dependent (DV1)	Creative Quality	Expert-assessed creativity score	Adapted CAT (4 dimensions, 5-point scale)
Dependent (DV2)	Efficiency	Total task completion time (hours)	Daily time-tracking logs
Dependent (DV3)	User Engagement	Audience rating + recall accuracy	Survey (3 items) + Delayed recall test
Covariate	Design Aptitude	Pre-experiment design skill equivalence	Standardized aptitude test + GPA

Note. CG = Control Group (n = 15); EG = Experimental Group (n = 15). CAT = Consensual Assessment Technique.

4.7 Data Analysis

All quantitative analyses were performed using IBM SPSS Statistics 27. Independent-samples t-tests were used for group comparisons on all continuous outcome measures; the significance threshold was set at $\alpha = .05$. Effect sizes were computed using Cohen's d (Cohen, 1988): values of .20, .50, and .80 indicate small, medium, and large effects, respectively. Chi-square tests were used for the categorical recall accuracy comparison. Qualitative thematic analysis followed Braun and Clarke's (2006) six-phase framework; themes were developed inductively and subsequently mapped onto HACE Model propositions in the integration phase.

V. RESULTS

5.1 Creative Quality (DV1): Expert Evaluation Findings

H1 was supported. Experimental Group designs achieved statistically significantly higher scores than Control Group designs on all four CAT dimensions and on the overall mean creativity score. The overall difference (EG: M = 3.87, SD = 0.43; CG: M = 3.21, SD = 0.51) corresponded to $t(28) = 3.84$, $p = .001$, $d = 1.40$ — a large effect size indicating a practically meaningful and clearly perceptible superiority in assessed creative quality.

The largest dimension-specific difference was observed on Conceptual Strength ($d = 2.14$), followed by Originality ($d = 1.61$), Aesthetic Quality ($d = 1.16$), and Relevance to Brief ($d = .77$). The smaller effect on Relevance to Brief is theoretically consistent with HACE Proposition 1's qualification that AI amplifies human creativity without substituting for the designer's judgment about contextual appropriateness a judgment that is not systematically enhanced by AI's generative capacity and that requires domain expertise and cultural knowledge that AI tools do not currently provide.

Table. Creative Quality Scores by Dimension — Control Group vs. Experimental Group

CAT Dimension	CG M (SD)	EG M (SD)	Mean Diff.	t(28)	p / d
Originality	3.15 (0.52)	3.91 (0.41)	+0.76	4.41	.002 / 1.61
Aesthetic Quality	3.33 (0.49)	3.88 (0.44)	+0.55	3.19	.009 / 1.16
Relevance to Brief	3.47 (0.44)	3.79 (0.39)	+0.32	2.10	.041 / .77
Conceptual Strength	2.89 (0.57)	3.90 (0.42)	+1.01	5.54	< .001 / 2.14
Overall Mean	3.21 (0.51)	3.87 (0.43)	+0.66	3.84	.001 / 1.40

Note. N = 30 (CG n = 15, EG n = 15). M = Mean; SD = Standard Deviation; Mean Diff. = EG minus CG; d = Cohen's d. All dimensions scored on 5-point Likert scale. One-tailed significance values reported for directional hypotheses.

A noteworthy within-group finding was that three EG participants produced designs rated below their pre-experiment baseline on the Originality dimension by at least one evaluator. Qualitative interview data identified these participants as experiencing creative paralysis by choice consistent with HACE Proposition 4's prediction of system-level quality degradation when AI assistance overwhelms rather than amplifies human creative direction. These cases are discussed in detail in Section 6.2.

5.2 Efficiency (DV2): Time-Tracking Analysis

H2 was supported. The Experimental Group completed all four tasks in a mean total of 31.4 hours (SD = 4.2) compared to 47.6 hours for the Control Group (SD = 5.8): $t(28) = 8.71$, $p < .001$, $d = 3.18$. This exceptionally large effect size reflects the magnitude of AI's impact on design workflow productivity. The overall time reduction of 34% is consistent with Adobe's (2023) industry-survey figure of 37% and with McKinsey's (2023) sector analysis, lending external validity to the experimental finding.

Phase-level analysis confirmed H2's specific prediction that efficiency gains would be greatest in the ideation phase. Ideation showed a 52%-time reduction for the EG (EG: M = 7.1 hrs; CG: M = 14.8 hrs), more than double the proportional reduction in any other phase. The final review phase showed no significant difference (p = .38), consistent with the theoretical prediction that AI does not reduce the cognitive load of high-judgment evaluative tasks only the mechanical production demands of generative tasks.

Table. Task Completion Time by Phase — Control Group vs. Experimental Group

Workflow Phase	CG M (SD) hrs	EG M (SD) hrs	Difference (hrs)	% Reduction	p
Ideation	14.8 (3.1)	7.1 (1.8)	−7.7	52%	< .001
Initial Drafting	12.3 (2.4)	9.3 (2.0)	−3.0	24%	.002
Refinement	13.8 (2.9)	8.8 (2.1)	−5.0	36%	< .001
Final Review	6.7 (1.8)	6.2 (1.6)	−0.5	7%	.38 (ns)
Total	47.6 (5.8)	31.4 (4.2)	−16.2	34%	< .001

Note. ns = not statistically significant. Peer-feedback integration time is included within the Refinement phase. Differences calculated as EG mean minus CG mean.

The non-significant difference in final review time is particularly theoretically meaningful: it suggests that the cognitive work of final quality judgment — deciding that a design is complete, coherent, and ready — is not accelerated by AI assistance. This finding is consistent with HACE Proposition 2's claim that AI reduces extraneous load (production demands) rather than intrinsic load (the cognitive complexity of creative judgment itself) and supports the model's characterization of the human designer as an irreplaceable evaluative and directorial agent in the collaborative process.

5.3 User Engagement (DV3): Audience Survey and Recall Results

H3 was supported. Experimental Group designs achieved significantly higher ratings than Control Group designs on all three engagement dimensions and on the composite engagement score. Large effect sizes were observed for Attention Capture (d = .82) and Emotional Appeal (d = .80), with a medium-to-large effect for Likelihood of Behavioural Response (d = .68). The composite engagement difference (EG: M = 4.02, SD = 0.67; CG: M = 3.46, SD = 0.78; t(78) = 3.97, p < .001, d = .73) was both statistically and practically significant.

Table. User Engagement Ratings — Control Group vs. Experimental Group Designs

Engagement Dimension	CG M (SD)	EG M (SD)	Mean Diff.	t(78)	p / d
Attention Capture	3.56 (0.74)	4.12 (0.61)	+0.56	4.29	< .001 / .82
Emotional Appeal	3.44 (0.79)	4.03 (0.68)	+0.59	4.17	< .001 / .80
Likelihood of Response	3.38 (0.82)	3.91 (0.71)	+0.53	3.56	.003 / .68
Composite Score	3.46 (0.78)	4.02 (0.67)	+0.56	3.97	< .001 / .73

Note. n = 40 respondents; each respondent rated 12 designs (6 CG, 6 EG). Composite = mean of three dimensions; Cronbach's α = .83.

The delayed recall assessment produced a 19.2 percentage-point difference in recognition accuracy: EG designs were correctly recalled by 67.3% of respondents compared to 48.1% for CG designs ($\chi^2(1) = 14.7$, $p < .001$). Element-level recall specificity the ability of respondents to accurately describe a specific visual element from a recalled design was also significantly higher for EG designs (EG: 58.4%; CG: 39.7%; $\chi^2(1) = 11.2$, $p < .001$). Critically, none of the 40 respondents correctly identified which designs were AI-assisted when asked directly, indicating that Human–AI collaborative designs were not perceived as generically or mechanically produced.

5.4 Task-Level Analysis

Task-level disaggregation of creativity scores revealed differential patterns that are theoretically informative. Task 1 (Social Media Campaign) showed the largest overall creativity gain (EG: M = 3.94; CG: M = 3.08; $p < .001$), consistent with AI's particular advantage in rapid multi-asset generation. Task 2 (Cultural Festival Poster) showed the smallest Relevance to Brief difference (EG: M = 3.71; CG: M = 3.52; $p = .19$, not significant), suggesting that the cultural specificity requirements of the brief constrained AI's contribution to design appropriateness a finding directly consistent with HACE Proposition 1's qualification about cultural knowledge and contextual judgment.

Task 3 (Product Packaging) produced the strongest improvement on Aesthetic Quality (EG: M = 4.01; CG: M = 3.22; $p < .001$), while Task 4 (Infographic) showed the strongest improvement on Originality (EG: M = 3.88; CG: M = 3.11; $p = .004$) but more modest gains on Relevance to Brief (EG: M = 3.86; CG: M = 3.41; $p = .038$), reflecting the challenges of applying generative image AI to data visualization contexts.

Table. Task-Specific Mean Creativity Scores (Overall CAT) Control vs. Experimental Group

Task	CG Mean	EG Mean	Difference	p-value
T1: Social Media Campaign	3.08	3.94	+0.86	< .001
T2: Cultural Festival Poster	3.31	3.71	+0.40	.048
T3: Product Packaging	3.18	3.98	+0.80	< .001
T4: Public Health Infographic	3.27	3.85	+0.58	.004
Grand Mean	3.21	3.87	+0.66	.001

Note. All mean scores on 5-point scale. Task 2 Relevance to Brief dimension not significant (p = .19) at the sub-dimension level despite overall significant task difference.

5.5 Qualitative Findings: Thematic Analysis

Thematic analysis of the 15 post-experiment semi-structured interviews with EG participants produced four primary themes, each with theoretical relevance to the HACE Model.

Theme 1 — Role Redefinition (Maker to Director): Twelve of fifteen EG participants described a fundamental shift in how they understood their own creative role from being the primary generator of visual material to functioning as a creative director who articulated vision, guided AI output, and exercised editorial judgment about what to select, combine, and refine. This role shift corresponds directly to HACE Proposition 1's characterization of H as the directorial component of the H-A relationship. The three participants who experienced the role shift negatively (perceiving it as a loss of creative ownership) are the same three who exhibited creative paralysis by choice in the quantitative data.

Theme 2 — Confidence Through Iterative Exploration: Most EG participants described a significant increase in creative risk-taking, attributed to the reduced psychological cost of abandoning a direction when alternatives can be generated quickly. This finding is theoretically interpretable through HACE Proposition 2: by reducing the cognitive investment cost of initial visual production, AI assistance lowers the commitment threshold for generative exploration, enabling designers to pursue more adventurous conceptual directions without the fear of sunk-cost abandonment.

Theme 3 — Creative Paralysis by Choice: Three participants reported a consistent experience of cognitive overwhelm when confronted with high volumes of AI-generated alternatives, resulting in difficulty maintaining creative direction and making decisive aesthetic choices. This pattern is a direct empirical instantiation of HACE Proposition 4 the model's prediction that system-level creative quality degrades when A overwhelms H. It also aligns with Schwartz's (2004) paradox of choice, and with Inie et al.'s (2023) empirical finding of decision paralysis under conditions of AI-generated abundance.

Theme 4 — Prompt Engineering as Acquired Creative Skill: All fifteen EG participants identified the ability to craft effective text prompts as a significant and learnable skill that improved substantially over the six-week period. Initial prompts were typically described as generic and literal; over time, participants developed more sophisticated prompting strategies incorporating mood, cultural reference, compositional principle, and stylistic affinity. This skill acquisition mirrors the development of other design communication skills the ability to specify a creative vision precisely enough that another agent (whether human collaborator or AI system) can productively respond to it.

VI. DISCUSSION

6.1 The HACE Model: Empirical Support and Theoretical Elaboration

The experimental findings provide consistent and multi-dimensional support for the HACE Model's four core propositions. HACE Proposition 1 that AI assistance amplifies rather than substitutes for human creative output is supported by the creativity findings: EG designs achieved higher creative quality not despite human involvement but because of it. The AI-generated elements were consistently refined, curated, and integrated through deliberate human directorial judgment, producing designs that expert evaluators found more conceptually coherent and more visually original than those produced without AI support.

The largest effect size was observed on Conceptual Strength ($d = 2.14$), the CAT dimension most directly reflective of the human designer's ability to unify disparate visual elements around a coherent creative idea. This is precisely the dimension that should be most improved by HACE Proposition 1's predicted amplification effect: the AI provides generative abundance, but the conceptual integration that gives the design its coherence and clarity is a fundamentally human cognitive achievement. The fact that this is where the largest quality gain occurred and not on the more mechanically operationalized Aesthetic Quality dimension is a theoretically meaningful finding.

HACE Propositions 2 and 3, concerning the cognitive load mechanism of efficiency improvement, are supported by the phase-level time analysis. The concentration of efficiency gains in the ideation phase (52% reduction) and the absence of significant efficiency gains in the final review phase are precisely consistent with the theoretical prediction that AI reduces extraneous load (production demands) rather than intrinsic load (the complexity of creative judgment). If AI were simply making designers work faster across the board, we would expect proportional time reductions across all phases; the finding of a highly uneven distribution, with the largest gains where production demands are highest and no gain where judgment demands are highest, is the signature of a cognitive load offloading mechanism rather than a general acceleration effect.

HACE Proposition 4, predicting system-level quality degradation when A overwhelms H, is corroborated by the creative paralysis cases. These three participants are not outliers or failures of the experimental design; they are theoretically predicted by the model as instances in which the human directorial component (H) was insufficient to manage the generative abundance provided by the AI component (A), resulting in the Schwartz (2004) choice-overload degradation of creative outcome quality. Their presence in the data strengthens the model's explanatory range by demonstrating that its predictions apply to negative as well as positive outcomes.

6.2 Human Creativity Amplification and the Director's Shift

The qualitative theme of role redefinition from maker to director has profound implications for how we understand creative identity in the age of intelligent tools. The majority of EG participants experienced this shift positively, describing it as a liberation from the mechanical demands of visual production and an opportunity to engage more fully with the conceptual and communicative dimensions of design. This experience aligns with what Shneiderman (2022) terms the human-centered AI ideal: AI that expands the range of human creative choices while ensuring that the human remains the author of the decisive creative acts.

The three participants who experienced the role shift negatively present a more complex picture. Their resistance to the director's role may reflect a conception of design craft in which the physical and cognitive act of making is intrinsically valuable not merely as a means to a designed output but as a form of personal expression and skill development. This conception is not indefensible; there are strong arguments in the broader philosophy of craft that the process of skilled making is constitutive of creative identity in ways that cannot be reduced to the quality of the outcome (Sennett, 2008). Design education that integrates AI tools must be sensitive to this dimension and must find pedagogical frameworks that honour the value of craft while equipping students for a professional landscape in which AI collaboration is increasingly standard.

6.3 Cultural Specificity and the Limits of AI Generative Knowledge

The task-level finding that EG designs showed the smallest creativity advantage on the Cultural Festival Poster task and specifically that the Relevance to Brief advantage was not statistically significant for this task is one of the most practically important findings of this study. It directly instantiates the HACE Model's qualification in Proposition 1 that AI assistance amplifies existing human capacity but does not substitute for domain-specific cultural knowledge and contextual judgment.

Current generative AI models are trained predominantly on corpora that over-represent Western visual culture and under-represent the visual traditions of South Asia, Africa, East Asia, and other non-Western regions. The consequence for Indian designers is that AI tools default to generic global aesthetic conventions unless actively and skillfully guided toward culturally specific references — a guidance that requires both deep knowledge of Indian visual traditions and advanced prompt engineering skills that the EG participants had not yet fully developed within the six-week study period.

This finding has direct implications for the design of AI tools and training data. If generative AI systems are to serve as genuine creative resources for designers working in diverse cultural contexts, their training corpora must be substantially diversified to represent the full range of human visual culture. Absent this diversification, AI collaboration risks functioning as a force for aesthetic globalization a systematic pressure toward the visual conventions that dominate Western digital media at the expense of the culturally specific visual languages that give regional design practice its distinctiveness and authenticity.

6.4 User Engagement and the Memory Advantage

The finding that audiences recalled specific visual elements from EG designs at rates nearly 20 percentage points higher than for CG designs, one week after initial exposure, is potentially one of the most commercially and communicatively significant results of this study. In real-world design practice where the effectiveness of a marketing campaign, a public health message, or a cultural communication is often measured in terms of whether the audience remembers what they saw this memory advantage translates directly into practical value.

The theoretical interpretation, consistent with the Von Restorff effect and with the aesthetic-usability literature (Tractinsky et al., 2000), is that the greater visual distinctiveness of EG designs produced by access to a broader generative vocabulary and by the compositional boldness associated with reduced production pressure creates stronger mnemonic encoding during initial exposure. Designs that break from visual convention in deliberate and aesthetically coherent ways are both more attention-capturing at first exposure and more memorable over time.

The finding that no respondent correctly identified AI-assisted designs as such is significant from an attribution perspective. It suggests that, in the context of Human–AI collaboration guided by a skilled human designer, the AI's contribution is fully integrated into a coherent aesthetic whole rather than remaining perceptible as a stylistically distinct foreign element. This integration is the hallmark of successful creative collaboration whether between two human designers or between a human designer and an intelligent tool and it suggests that the attribution effect documented by Kim et al. (2022) is not an inherent feature of AI-assisted design but a consequence of AI use that is insufficiently integrated with human creative judgment.

6.5 Ethical Dimensions: Attribution, Disclosure, and Labor

The ethical questions raised by AI-assisted design cannot be separated from the empirical questions, because the value of the empirical findings depends in part on the ethical framework within which they are interpreted. This paper adopts three positions on the central ethical questions.

On authorship: Human–AI collaborative design constitutes genuine human creative authorship, mediated by an intelligent tool, in the same way that photography constitutes genuine human creative authorship mediated by a camera. The presence of the AI tool does not diminish the human creative act; it changes the nature of the skill required to perform it from technical production skill toward directorial judgment, conceptual synthesis, and aesthetic curation. The designer who effectively directs and curates AI output to produce a high-quality design has exercised genuine creative agency.

On disclosure: Designers and studios using AI tools in professional work have an obligation to be transparent about this use in their client relationships, even if audiences cannot identify AI-assisted designs as such from the outputs alone. Transparency is owed not because AI assistance diminishes the work's quality the evidence suggests it often improves it but because clients and audiences have a legitimate interest in understanding how creative work they commission or receive was produced.

On labor: The efficiency gains documented in this study a 34% reduction in task completion time represent a genuine productivity improvement, but their economic implications depend entirely on how the saved time is reallocated. If AI efficiency gains are used to reduce designer employment rather than to enable designers to take on more complex and rewarding work, the net social effect could be negative despite the output quality improvements. Design organizations, professional bodies, and institutions have a responsibility to advocate for productivity-sharing arrangements that distribute AI efficiency gains fairly across all stakeholders in the design value chain.

VII. RESEARCH CONTRIBUTIONS

This study makes seven distinct contributions to the literature on Human–AI collaboration, design research, and creative education:

Contribution 1: The HACE Model

The most significant theoretical contribution of this paper is the HACE Model the first conceptual framework to specifically integrate Cognitive Load Theory, creativity theory, and Human–AI collaboration theory for the context of graphic design. The model generates testable predictions about the conditions under which Human–AI collaboration improves, maintains, or degrades creative quality, provides a theoretical account of the cognitive mechanism underlying efficiency improvements, and identifies the critical role of human directorial judgment as the limiting factor in collaborative creative systems. The HACE Model is offered as a replicable theoretical instrument for future research across design domains and populations.

Contribution 2: Experimental Evidence with Validated Instruments

This study provides controlled experimental evidence as opposed to the retrospective surveys and industry reports that dominate the existing literature on the effects of Human–AI collaborative design on creative quality, efficiency, and user engagement. The use of the validated Consensual Assessment Technique (Amabile, 1983), prospective time-tracking, and a two-stage audience evaluation protocol with delayed recall represents a methodological standard not previously applied to this research question in the Indian design education context.

Contribution 3: The Non-Western Design Education Perspective

By conducting this research in an Indian university context, with design tasks that include culturally specific requirements (bilingual poster design, artisanal Indian product packaging), this study contributes a non-Western perspective to a literature dominated by North American and European research. The finding that AI's advantage is attenuated in culturally specific design contexts due to training data biases is a contribution that has direct practical implications for the global deployment of AI design tools.

Contribution 4: User Engagement and Memory Evidence
The delayed recall findings showing a 19.2 percentage-point advantage for Human–AI collaborative designs over traditionally produced designs, one week after exposure provide the first controlled experimental evidence of a memory advantage for AI-assisted design in the published literature. This finding has substantial practical relevance for design effectiveness evaluation and contributes to the user engagement literature by extending the concept of AI-assisted quality improvement from expert assessment to lay audience response.

Contribution 5: Creative Paralysis by Choice
The identification and theorization of creative paralysis by choice the system-level degradation of creative quality when AI generative abundance exceeds the human designer's capacity for decisive aesthetic judgment contributes a novel construct to the literature on Human–AI collaboration. This construct bridges Schwartz's (2004) paradox of choice, Sweller's (1988) cognitive load framework, and Inie et al.'s (2023) empirical findings, providing a theoretically grounded account of the conditions under which AI collaboration produces negative rather than positive creative outcomes.

Contribution 6: Prompt Engineering as Creative Literacy
The qualitative finding that prompt engineering the skill of specifying a creative vision precisely enough for an AI system to generate productive responses is an acquired creative competency that improves with deliberate practice contributes to the emerging literature on AI literacy in design education. The characterization of prompt engineering as a creative skill analogous to other forms of visual specification (briefing, storyboarding, style guidelines) provides a pedagogically actionable framework for its incorporation into design curricula.

Contribution 7: Phase-Level Efficiency Analysis
The phase-level decomposition of efficiency gains demonstrating that AI produces a 52% time reduction in ideation but no significant reduction in final review provides a theoretically and practically more informative account of AI's workflow impact than the aggregate efficiency figures reported in industry surveys. This phase-level analysis confirms the HACE Model's cognitive load mechanism, distinguishes between AI's effect on production demands and its non-effect on judgment demands, and provides specific guidance for design workflow redesign in AI-integrated studios and educational programs.

VIII. CONCLUSIONS AND RECOMMENDATIONS

8.1 Summary of Principal Findings
This study set out to provide rigorous experimental evidence about what happens when graphic design students work with artificial intelligence as a creative partner, and to develop a theoretical framework capable of explaining the mechanisms behind the observed outcomes. The principal findings are as follows.

First, Human–AI collaborative design produces work of measurably higher creative quality than traditional unaided design, across all four dimensions of a validated expert assessment instrument, with large effect sizes ($d = .77$ to 2.14) indicating practically meaningful rather than merely statistically detectable differences. The largest advantage is on Conceptual Strength the dimension most reflective of the human designer's integrative creative judgment supporting the HACE Model's characterization of AI assistance as an amplifier of human creative capacity rather than a substitute for it.

Second, AI tool integration reduces total task completion time by approximately 34%, with the greatest proportional savings in the ideation phase (52% reduction). The absence of significant efficiency gains in the final review phase is consistent with the HACE Model's prediction that AI reduces extraneous cognitive load (production demands) but not intrinsic cognitive load (the complexity of final creative judgment). Efficiency gains are largest precisely where cognitive load theory predicts they should be.

Third, Human–AI collaborative designs generate stronger audience engagement responses across all measured dimensions, with a particularly notable 19.2 percentage-point advantage in delayed recall accuracy evidence that AI-assisted designs create stronger mnemonic traces in audience memory, with significant implications for design effectiveness in real-world communication contexts.

Fourth, the creative paralysis by choice phenomenon experienced by three of fifteen EG participants is both an empirical finding and a theoretical prediction of the HACE Model's fourth proposition. It demonstrates that AI collaboration is not uniformly beneficial and identifies the quality of human directorial judgment as the critical variable determining whether Human–AI collaboration enhances or degrades creative outcomes.

8.2 Recommendations

For Design Educators

- Incorporate AI tool literacy including prompt engineering, AI output curation, and workflow integration as a core professional competency in design curricula from the first year, equivalent in pedagogical status to typography, color theory, and visual composition.
- Develop explicit assessments of aesthetic judgment under conditions of generative abundance — assignments that require students to select, justify, and refine from a large pool of AI-generated options rather than simply generate outputs from scratch.
- Create dedicated modules addressing the cultural specificity challenges of AI-assisted design in non-Western contexts, equipping Indian students with the prompt engineering strategies needed to guide AI tools toward culturally authentic visual references.
- Reframe assessment criteria to recognize and reward the creative director role: evaluate students on the quality of their directorial decisions and final integrated outcomes, not merely on technical production skill.

For Professional Design Studios and Practitioners

- Invest in structured AI tool training for all design staff, recognizing that effective Human–AI collaboration is a skill that requires deliberate development rather than a capability that emerges automatically from tool access.
- Develop clear internal ethical guidelines for AI-assisted work, addressing client disclosure norms, attribution standards, and the cultural appropriateness review requirements for AI-generated visual content.
- Redesign workflow structures to leverage AI's efficiency gains in ideation and drafting while preserving the uncompressed time for final quality judgment that the evidence shows AI does not accelerate.

For AI Tool Developers

- Invest in training corpora that adequately represent non-Western visual traditions, enabling AI tools to serve as genuine creative resources for designers working in diverse cultural contexts rather than defaulting to Western aesthetic conventions.
- Design interface features that support decisive judgment under abundance tools for organizing, comparing, and progressively narrowing AI-generated option sets rather than maximizing raw generative output volume.

For Researchers

- Conduct longitudinal studies tracking the creative development of designers who work extensively with AI tools over multi-year periods, to determine whether Human–AI collaboration produces durable enhancement or gradual erosion of specific creative competencies.
- Replicate and extend the HACE Model across different design domains (UX, motion graphics, architectural visualization), different cultural contexts, and different levels of design expertise, to establish the model's boundary conditions and generalizability.
- Develop validated cognitive load measures specific to Human–AI design collaboration contexts, enabling direct testing of the HACE Model's proposed cognitive load mechanism rather than relying on efficiency time as a proxy.

8.3 Limitations

Several limitations of this study must be acknowledged. The sample is restricted to a single institution, a single city, and a single cultural context;

generalizability to other populations requires replication. The six-week duration, while sufficient to detect the experimental effects measured, does not capture longer-term dynamics of AI skill development or potential skill atrophy. The three AI platforms used represent a specific configuration of available tools in early 2026; the rapidly evolving landscape of generative AI means that findings may require revisiting as tool capabilities advance. Finally, the efficiency measure prospective time-tracking is subject to social desirability bias despite the verification protocols employed.

8.4 Concluding Remarks

The question of what artificial intelligence does to human creativity in design is not adequately answered by either utopian enthusiasm or dystopian alarm. The evidence of this study suggests a more nuanced and more interesting answer: AI, deployed as a collaborative partner guided by deliberate human creative direction, amplifies human creative capacity, accelerates the design process, and produces outputs that audiences find more engaging and more memorable. But this amplification is conditional it depends on the quality of the human directorial judgment brought to the collaboration, on the cultural intelligence applied to the curation of AI outputs, and on the metacognitive resilience to maintain creative direction under conditions of generative abundance.

The designer in the age of AI is not diminished; they are redefined. From maker to director. From technical executor to creative strategist. From sole author to collaborative intelligence. Understanding this redefinition clearly theoretically, empirically, and practically is among the most important tasks facing design research, design education, and the design profession in the years ahead. This study is offered as a contribution toward that understanding.

REFERENCES

[1] Adobe Inc. (2023). Adobe creative trends report 2023: The age of AI creativity. Adobe Systems Incorporated.

[2] Amabile, T. M. (1983). The social psychology of creativity: A consensual assessment technique. Journal of Personality and Social Psychology, 43(5), 997–1013. https://doi.org/10.1037/0022-3514.43.5.997

[3] Boden, M. A. (2004). The creative mind: Myths and mechanisms (2nd ed.). Routledge.

[4] Boyden, B. E. (2022). Copyright protection for AI-generated works: Lessons from photography and computer-generated works. Wisconsin Law Review, 2022(3), 601–648.

[5] Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

[6] Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research. Rand McNally.

[7] Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Lawrence Erlbaum Associates.

[8] Creswell, J. W., & Plano Clark, V. L. (2018). Designing and conducting mixed methods research (3rd ed.). SAGE Publications.

[9] Csikszentmihalyi, M. (1996). Creativity: Flow and the psychology of discovery and invention. HarperCollins.

[10] Dove, G., Halskov, K., Forlizzi, J., & Zimmerman, J. (2017). UX design innovation: Challenges for working with machine learning as a design material. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (pp. 278–288). ACM. https://doi.org/10.1145/3025453.3025739

[11] Elgammal, A., Liu, B., Elhoseiny, M., & Mazzone, M. (2017). CAN: Creative adversarial networks generating 'art' by learning about styles and deviating from style norms. arXiv:1706.07068. https://arxiv.org/abs/1706.07068

[12] European Union. (2024). Artificial Intelligence Act (Regulation (EU) 2024/1689). Official Journal of the European Union.

[13] Frich, J., MacDonald Vermeulen, L., Remy, C., Biskjaer, M. M., & Dalsgaard, P. (2019). Mapping the landscape of creativity support tools in HCI. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (pp. 1–18). ACM. https://doi.org/10.1145/3290605.3300619

[14] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets.

Advances in Neural Information Processing Systems, 27, 2672–2680.

[15] Guilford, J. P. (1967). The nature of human intelligence. McGraw-Hill.

[16] Hunt, R. R. (1995). The subtlety of distinctiveness: What von Restorff really did. Psychonomic Bulletin & Review, 2(1), 105–112. https://doi.org/10.3758/BF03214414

[17] Inie, N., Thomsen, J. K., & Dow, S. (2023). Designing AI-based creative tools: A practitioner perspective. In Proceedings of the 2023 ACM Designing Interactive Systems Conference (pp. 1218–1232). ACM. https://doi.org/10.1145/3563657.3595970

[18] Kantosalo, A., & Toivonen, H. (2016). Modes for creative human-computer collaboration. In Proceedings of the 7th International Conference on Computational Creativity (ICCC 2016) (pp. 77–84). Sony CSL.

[19] Karimi, S., Maher, M. L., Rezwana, J., & Begel, A. (2020). Creative ideation with and without AI assistance. In Proceedings of the Twelfth International Conference on Computational Creativity. Association for Computational Creativity.

[20] Kim, J., Giroux, M., & Lee, J. C. (2022). When do people think AI-generated art is art? Artworks and their artistic value attributed to AI. Computers in Human Behavior, 135, 107376. https://doi.org/10.1016/j.chb.2022.107376

[21] Kirsh, D. (2010). Thinking with external representations. AI & Society, 25(4), 441–454. https://doi.org/10.1007/s00146-010-0272-8

[22] Krippendorff, K. (2004). Reliability in content analysis: Some common misconceptions and recommendations. Human Communication Research, 30(3), 411–433. https://doi.org/10.1111/j.1468-2958.2004.tb00738.x

[23] Lubart, T. I. (2005). How can computers be partners in the creative process: Classification and commentary on the special issue. International Journal of Human-Computer Studies, 63(4–5), 365–369. https://doi.org/10.1016/j.ijhcs.2005.04.002

[24] McKinsey & Company. (2023). The state of AI in 2023: Generative AI's breakout year. McKinsey Global Institute.

[25] Paas, F. G. W. C., & van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. Journal of Educational Psychology, 86(1), 122–133. https://doi.org/10.1037/0022-0663.86.1.122

[26] Reinecke, K., & Bernstein, A. (2011). Improving performance, perceived usability, and aesthetics with culturally adaptive user interfaces. ACM Transactions on Computer-Human Interaction, 18(2), 1–29. https://doi.org/10.1145/1970378.1970382

[27] Rezwana, J., & Maher, M. L. (2023). Designing effective human-AI collaboration: The role of human and AI agency in co-creative systems. Proceedings of the 14th International Conference on Computational Creativity, 265–275.

[28] Schwartz, B. (2004). The paradox of choice: Why more is less. Ecco Press.

[29] Sennett, R. (2008). The craftsman. Yale University Press.

[30] Shneiderman, B. (2022). Human-centered AI. Oxford University Press.

[31] Sundar, S. S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. In M. J. Metzger & A. J. Flanagin (Eds.), Digital media, youth, and credibility (pp. 73–100). MIT Press.

[32] Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

[33] Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. Educational Psychology Review, 22(2), 123–138. https://doi.org/10.1007/s10648-010-9128-5

[34] Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. Educational Psychology Review, 10(3), 251–296. https://doi.org/10.1023/A:1022193728205

[35] Tractinsky, N., Katz, A. S., & Ikar, D. (2000). What is beautiful is usable. Interacting with Computers, 13(2), 127–145. https://doi.org/10.1016/S0953-5438(00)00031-X

[36] Yang, Q., Steinfeld, A., Rosé, C., & Zimmerman, J. (2023). Unremarkable AI: Fitting intelligent decision support into critical, clinical decision-making processes. In Proceedings of the 2023

CHI Conference on Human Factors in Computing Systems. ACM.

Appendix A: HACE Model Summary Table

The following table provides a concise reference summary of the HACE Model's components, theoretical derivations, and empirical operationalizations in the context of this study.

Table. HACE Model: Components, Theoretical Basis, and Empirical Operationalization

Component	Name	Theoretical Basis	Role in Model	Empirical Operationalization
H	Human Creativity	Amabile (1983); Csikszentmihalyi (1996); Guilford (1967)	Directorial and evaluative agent; the source of creative vision and contextual judgment	CAT scores (Originality, Conceptual Strength); qualitative themes 1 & 3
A	AI Assistance	Lubart (2005); Rezwana & Maher (2023); Elgammal et al. (2017)	Generative partner providing visual abundance and production velocity	AI tool access (Firefly, Midjourney, Canva AI); prompt engineering skill (theme 4)
C	Cognitive Load	Sweller (1988, 2010); Kirsh (2010); Paas & van Merriënboer (1994)	Mediating mechanism: A reduces extraneous C, freeing germane C for creative processing	Phase-level time distribution; qualitative theme 2
E	Efficiency	Goldschmidt (1991); Adobe (2023); McKinsey (2023)	Outcome metric: ratio of creative quality to time/cognitive resource investment	Total task completion time; phase-level breakdown (Table 2)

Appendix B: Complete Statistical Results Summary

Table. All Primary Outcome Measures — Statistical Results Summary

Outcome Measure	CG M	EG M	Diff.	t or χ^2	df	p / d
Overall Creativity (CAT)	3.21	3.87	+0.66	t = 3.84	28	.001 / 1.40
Originality	3.15	3.91	+0.76	t = 4.41	28	.002 / 1.61
Aesthetic Quality	3.33	3.88	+0.55	t = 3.19	28	.009 / 1.16
Relevance to Brief	3.47	3.79	+0.32	t = 2.10	28	.041 / .77
Conceptual Strength	2.89	3.90	+1.01	t = 5.54	28	< .001 / 2.14

Outcome Measure	CG M	EG M	Diff.	t or χ^2	df	p / d
Total Completion Time (hrs)	47.6	31.4	−16.2	t = 8.71	28	< .001 / 3.18
User Engagement (Composite)	3.46	4.02	+0.56	t = 3.97	78	< .001 / .73
Attention Capture	3.56	4.12	+0.56	t = 4.29	78	< .001 / .82
Emotional Appeal	3.44	4.03	+0.59	t = 4.17	78	< .001 / .80
Likelihood of Response	3.38	3.91	+0.53	t = 3.56	78	.003 / .68
Delayed Recall Rate (%)	48.1%	67.3%	+19.2%	χ^2 = 14.7	1	< .001 / —
Recall Specificity (%)	39.7%	58.4%	+18.7%	χ^2 = 11.2	1	< .001 / —

Note. CG = Control Group (n = 15); EG = Experimental Group (n = 15). For Engagement outcomes, df = 78 (40 respondents × 2 conditions). d = Cohen's d for t-tests; not reported for chi-square tests. All significance values are two-tailed except where directional hypotheses justified one-tailed testing.

— End of Research Paper —

Submitted to: Dr. Sukhada Khandge | School of Film & Media | ADYPU, Pune | April 2026