Type of the Paper (Article)

Cognitive Activation or Design Dependency? The Dual Mechanisms of Large Language Models in Shaping Individual Design Quality Across Tasks of Varying Complexity

TianCi Fang [1]*, ShunHao Liang [2] and Owen Powell [3]

[1] Wuhan Technology and Business University; 1106192773@qq.com
[2] Wuhan Technology and Business University; 544738762@qq.com
[3] Glyndwr University; djadrysava0517@outlook.com
* Correspondence: 1106192773@qq.com

Abstract

With the widespread integration of large language models (LLMs) in the field of design innovation, the mechanisms through which they influence individual designers' performance remain unclear. Existing studies primarily focus on the macro-level impact of LLMs on creative diversity, while systematic investigations into their micro-level effects on individual cognitive processes in specific design task contexts are lacking. Based on Cognitive Load Theory and Collaborative Cognitive Load Theory, this study proposes a dual-opposing-mechanism model to explain how LLMs affect individual design quality, introducing task complexity as a moderating variable to systematically examine the differential effects of LLMs across varying task contexts. In Experiment 1 (N = 216), a 2 (assistance type: LLM-assisted vs. human-assisted) × 2 (task complexity: simple vs. complex) between-subjects design was employed. Results indicated that for simple design tasks, LLM assistance significantly enhanced individual design quality through the activation of cognitive associations (cognitive activation pathway); conversely, for complex design tasks, LLM assistance led to a decrease in design quality due to induced design dependency (design dependency pathway). Experiment 2 (N = 216) further examined the effect of constraining LLM outputs, revealing that in complex tasks, constrained LLM responses effectively mitigated design dependency and improved design quality, whereas in simple tasks, such constraints weakened the cognitive activation effect, thereby reducing design quality. These findings elucidate the dual-opposing mechanisms of LLMs in design innovation contexts and provide theoretical

Academic Editor: Haiwen Wang

Received: July 21, 2025
Revised: August 22, 2025
Accepted: August 24, 2025
Published: September 28, 2025

foundations and practical guidance for differentiated deployment of LLM tools in design education and practice.

Keywords: large language models; design quality; cognitive activation; design dependency; task complexity; cognitive load theory

1. Introduction

Since the advent of ChatGPT, large language models (LLMs) have been widely applied across domains such as design, education, engineering, and business, profoundly transforming the way humans engage in creative work [1][2]. In the context of design innovation, LLMs have demonstrated strong capabilities in concept generation and solution expansion, positioning them as important tools for facilitating design innovation [3][4]. Designers and students increasingly rely on LLM-based tools such as ChatGPT and Midjourney to assist in the full spectrum of design tasks, from conceptual exploration to solution refinement [5]. However, whether the widespread integration of LLMs enhances design quality or subtly fosters cognitive dependency remains a subject of considerable debate [6][7].

Design quality is a core metric for evaluating the effectiveness of design innovation, encompassing multiple dimensions including novelty, feasibility, and aesthetic value [8]. Prior studies suggest that LLMs can enhance design outcomes by providing diverse design references and stimulating remote associative thinking [9]. Conversely, other research indicates that excessive reliance on LLM-generated solutions may undermine designers' independent thinking, resulting in a phenomenon termed "design dependency," in which designers perform incremental modifications within the framework provided by LLMs rather than generating original solutions independently [10][11]. This phenomenon is closely related to the concept of "creative fixation" in creativity research [12], yet its specific manifestation and underlying mechanisms within the context of design tasks remain underexplored.

Crucially, the complexity of design tasks may fundamentally alter the direction of LLMs' impact on design quality. In simple design tasks (e.g., visual element reorganization or single-function interface design), designers' cognitive load is relatively low, and the rich references provided by LLMs can facilitate cognitive activation and broaden design thinking [13]. In contrast, in complex design tasks (e.g., multi-system interaction design or user experience system planning), the inherent cognitive load is already high, and the structured outputs generated by LLMs may further occupy limited working memory resources, inducing cognitive overload and ultimately promoting design dependency [14][15]. The moderating role of task complexity in this context has not been systematically investigated.

This study aims to address these gaps by examining the following core questions: How do LLMs influence individual design quality through the opposing pathways of cognitive activation and design dependency? How does task complexity moderate the relative strength of these pathways? Can constraining LLM outputs effectively mitigate design dependency in complex tasks? Theoretically, this study extends Cognitive Load Theory to LLM-assisted design tasks, proposing a dual-opposing-mechanism model of LLM effects on design quality, and provides rigorous empirical evidence through two controlled experiments. Practically, the findings offer actionable guidance for educators and practitioners on differentiated deployment of LLM tools, with implications for optimizing human—AI collaborative design processes.

The structure of this paper is as follows: Section 2 reviews relevant literature on Cognitive Load Theory, LLMs, and design innovation, and presents the research hypotheses; Section 3 details the methodology of the two experiments; Section 4 reports the experimental results; Section 5 provides an in-depth discussion; and Section 6 concludes the study and outlines directions for future research.

2. Related Work

2.1. Cognitive Load Theory and Design Innovation

Cognitive Load Theory (CLT), proposed by Sweller in 1988, posits that human working memory has limited capacity and that three types of cognitive load exist during information processing: intrinsic load, extraneous load, and germane load [16]. Intrinsic load arises from the inherent complexity of the task, determined by the number of information elements and their interrelationships. Extraneous load stems from irrelevant or distracting information unrelated to the task's inherent complexity. Germane load is associated with learning and schema construction. When the sum of intrinsic and extraneous loads exceeds working memory capacity, cognitive overload occurs, thereby impeding higher-order thinking processes [17].

In the context of design innovation, CLT has been widely applied to explain the relationship between designers' cognitive processes and design quality [18]. Studies indicate that a moderate level of cognitive load can stimulate associative thinking, whereas excessive cognitive load inhibits the generation of original solutions [19]. Research by Cross et al. demonstrated that professional designers are able to maintain high design quality by effectively managing the intrinsic load of tasks through cognitive strategies [20]. However, when external information inputs—such as reference cases or design templates—are overly abundant, even experienced designers may experience cognitive overload, resulting in decreased originality of solutions [21].

Collaborative Cognitive Load Theory (CCLT) further extends the scope of CLT by examining cognitive load dynamics in collaborative contexts [22]. The theory suggests that collaboration can, on one hand, offload intrinsic cognitive load to free individual cognitive resources, but on the other hand, it may increase extraneous load due to communication costs and information interference [22]. This theoretical framework provides a crucial basis for understanding the dual effects of LLMs in collaborative design settings.

2.2. The Impact of LLMs on Design Innovation

In recent years, research on the application of LLMs in the design domain has grown rapidly. On one hand, multiple studies have demonstrated the positive effects of LLMs on design innovation. Eapen et al. found that LLMs can facilitate cross-domain conceptual associations, helping designers break cognitive fixedness and generate more innovative design solutions [23]. Boussioux et al. reported that in creative problem-solving tasks, human—AI collaboration outperforms either purely human collaboration or AI working independently [24]. In the context of design education, Urban et al. demonstrated that ChatGPT can significantly enhance university students' performance on design-related creative tasks [25].

On the other hand, the potential negative effects of LLMs on design innovation should not be overlooked. Doshi and Hauser observed that while LLMs can improve individual creative performance, they may simultaneously lead to solution homogenization at the group level [26]. Anderson et al. noted that long-term use of LLMs may degrade designers' independent thinking abilities, fostering over-reliance on AI-generated content [27]. In the study of design fixation, Crilly's review highlighted that excessive exposure to external reference cases is a major contributor to design fixation [12]; LLMs, with their capability to rapidly generate large volumes of structured solutions, may therefore pose a potential risk for inducing design fixation.

Notably, existing research on the effects of LLMs on design quality has often overlooked task complexity as a critical moderating variable. While Lee and Chung identified a significant effect of ChatGPT on creativity, they did not differentiate the effects across tasks of varying complexity [28]. Similarly, Rafner et al., in discussing creativity in the era of generative AI, emphasized that future research should examine how task characteristics moderate the effectiveness of AI assistance [29]. The present study directly responds to this call by systematically investigating the moderating role of task complexity in LLM-assisted design.

2.3. Design Task Complexity and Cognitive Processes

Design task complexity is a key factor influencing both the design process and design quality [30]. Kirschner et al. operationalized task complexity in terms of the number of interacting information elements within a task, positing that the greater the number of elements and the stronger their interconnections, the higher the intrinsic cognitive load of the task [31]. In the design domain, simple design tasks (e.g., creative recombination of a single visual element) typically involve fewer interacting elements, whereas complex design tasks (e.g., overall planning of multi-system interactive interfaces) entail numerous interdependent design constraints and user requirements, resulting in significantly higher intrinsic cognitive load [32].

Task complexity not only directly affects design quality but also indirectly influences the effectiveness of external information processing by moderating the allocation of cognitive resources [33]. In low-complexity tasks, designers' working memory retains sufficient residual capacity to effectively integrate external information and transform it into innovative solutions [34]. In high-complexity tasks, however, working memory is largely consumed by the intrinsic load of the task itself, and additional external information is more likely to trigger cognitive overload. This, in turn, leads designers to adopt cognitive shortcut strategies, such as directly utilizing external solutions rather than engaging in independent thinking [35]. This mechanism provides critical theoretical support for the core hypotheses of the present study.

2.4. Research Hypotheses

Based on the theoretical analysis above, the study proposes the following hypotheses:

- H1: Compared to human-assisted support, LLM assistance significantly enhances individual design quality in simple design tasks, and this effect is positively mediated by cognitive activation;
- H2: Compared to human-assisted support, LLM assistance significantly reduces individual design quality in complex design tasks, and this effect is negatively mediated by design dependency;
- H3: Task complexity moderates the effect of LLM assistance on design quality, such that the cognitive activation pathway is more pronounced in simple tasks, whereas the design dependency pathway is more pronounced in complex tasks;
- H4): In complex design tasks, constrained-response LLMs (as opposed to batch-response LLMs) improve design quality by reducing design dependency; in simple design tasks, constrained-response LLMs reduce design quality by weakening the cognitive activation effect.

3. Methods

3.1. Research Strategy

This study employed an experimental research approach, using two independent controlled experiments to systematically test the proposed hypotheses. Experiment 1 was designed to examine the dual-opposing mechanisms through which LLMs influence design quality and the moderating role of task complexity (H1, H2, H3). Building on Experiment 1, Experiment 2 further investigated the intervention effect of constraining LLM outputs (H4). Both experiments employed a between-subjects design to eliminate potential learning and order effects.

The overall methodological framework followed a logical progression of “theoretical modeling → manipulation testing → mediation analysis → intervention verification” (see Figure 1).

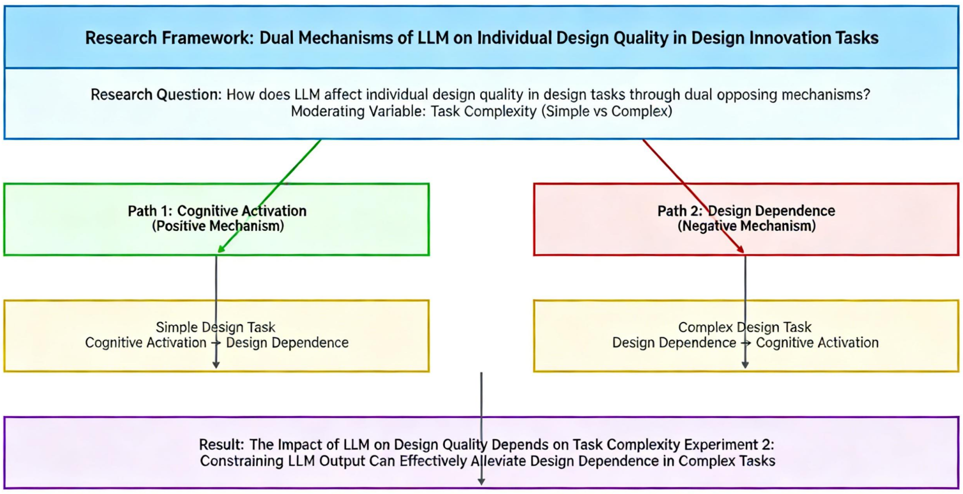

Figure 1. Schematic diagram of the research framework, illustrating the dual-opposing mechanisms of LLMs on design quality and the moderating role of task complexity.

3.2. Experimental Materials and System

The experiments were conducted on a self-developed online human—AI collaborative design platform. The platform integrated multiple functional modules, including task instructions, a timer, a design input interface, a collaborator dialogue box, and a solution display area. In the LLM-assisted conditions, the platform incorporated a GPT-4-based design assistant model, with response strategies configured according to the experimental conditions.

Task Design: Based on the operationalization of task complexity in Cognitive Load Theory, two types of design tasks were developed. The simple design task required participants to propose a creative exterior design for a common daily item (a portable power bank), involving relatively few design constraints and limited interacting information elements. The complex design task required participants to design a complete user interaction interface for a smart home control system for a company, encompassing multiple user scenarios, various device types, and complex interaction logic, thereby involving a significantly higher number of interacting information elements.

LLM-Assisted Content: To control the quality of LLM assistance, the research team pre-generated 100 design suggestions for each task type using GPT-4. Two independent experts rated the creativity of these suggestions on a 5-point scale. The average creativity score for the simple task pool was 2.18 (SD = 0.39), and for the complex task pool, 2.21 (SD = 0.41); the difference was not significant ($t_{98} = 0.41$, $p = 0.68$), ensuring baseline consistency of the quality of LLM-assisted content.

Human-Assisted Content: In the human-assisted conditions, trained research assistants acted as collaborators, providing design suggestions drawn from a pre-prepared human solution pool. This pool consisted of 50 solutions per task type contributed by eight design-major students. The average creativity scores of the human solution pool (simple task: M = 2.20, SD = 0.40; complex task: M = 2.18, SD = 0.39) did not differ significantly from the LLM solution pool, thereby eliminating content quality differences as a confounding factor.

3.3. Experiment 1

Participants: A total of 216 participants were recruited through an online platform. All participants had at least six months of experience using LLMs and possessed basic design knowledge. The mean age was 22.4 years (SD = 1.8), with 44.4% male and 55.6% female; 91.2% held a bachelor's degree or higher. Participants were randomly assigned to one of four experimental conditions, with 54 participants per group: human-assisted/simple task, LLM-assisted/simple task, human-assisted/complex task, and LLM-assisted/complex task.

Procedure: Participants first completed a demographic questionnaire and then logged into the experimental platform to perform the assigned design task within 10 minutes. During the task, participants could view design suggestions provided by the collaborator (human or LLM) at any time and submit their own designs through the text input interface. Upon task completion, participants completed a manipulation check questionnaire and additional measurement scales.

Measures: Design quality was assessed using a dual-measurement strategy, incorporating both subjective self-assessment and objective expert ratings. Subjective design quality was measured with a single-item scale ("Please rate your design performance in this task," 5-point scale). Objective design quality was independently evaluated by two trained experts who were blind to the experimental conditions. They rated each participant's design on novelty and feasibility (5-point scale). Inter-rater reliability was high (ICC = 0.853), and the mean of the two ratings was used as the final objective design quality score.

Cognitive activation was measured using an adapted inspiration scale from Böttger et al. (5 items, e.g., "My thinking was fully activated during the design process," α = 0.961) [36]. Design dependency was assessed based on the creative fixation measurement method proposed by Lu et al. [37]. Two experts independently evaluated the similarity between participants' designs and the assisted content using a 3-point scale (0 = no similarity, 1 = partial similarity, 2 = high similarity). Inter-rater reliability was high (ICC = 0.879), and the average score was used as the design dependency measure.

3.4. Experiment 2

Participants: Another 216 participants were recruited, with the same eligibility criteria as in Experiment 1. The mean age was 23.1 years (SD = 2.1), with 41.2% male and 58.8% female; 93.5% held a bachelor's degree or higher. Participants were randomly assigned to one of four experimental conditions, with 54 participants per group: batch-response LLM/simple task, constrained-response LLM/simple task, batch-response LLM/complex task, and constrained-response LLM/complex task.

LLM Response Type Manipulation: In the batch-response condition, each LLM output contained 10 different design suggestions. In the constrained-response condition, each LLM output contained only 2 design suggestions. Both response types drew from the same solution pool; the only difference was the number of suggestions provided per response. The response interval was set to 100 seconds for both conditions to ensure that participants had sufficient time to process the assisted information.

Measures: The same dual-measurement strategy as in Experiment 1 was employed to assess design quality, along with measurements of cognitive activation and design dependency.

3.5. Data Analysis

The following statistical analysis methods were employed in this study: (1) Manipulation check: Independent-samples t-tests were conducted to verify the effectiveness of the task complexity manipulation; (2) Main and interaction effects: A 2 × 2 between-subjects analysis of variance (ANOVA) was used to examine the effects of assistance type, task complexity, and their interaction on design quality; (3) Moderated mediation analysis: PROCESS macro (Model 8) was applied to test moderated mediation effects, with 10,000 bootstrap resamples used to estimate confidence intervals for indirect effects [38]; (4) Sensitivity analysis: All analyses were repeated after excluding three participants with missing data due to equipment failure, in order to assess the robustness of the results. All statistical analyses were conducted using SPSS 26.0 and R 4.2.0.

4. Data

4.1. Basic Data Information

TA total of 216 participants were recruited for Experiment 1, of which 3 cases had missing design quality scores due to equipment failure (missing rate = 1.39%), resulting in a final valid sample of 213 participants. Similarly, Experiment 2 recruited 216 participants, with 3 cases of missing data (missing rate = 1.39%), yielding a valid sample of 213 participants. No significant differences were observed between the participants in the two experiments in terms of baseline characteristics, including gender, age, educational level, and prior LLM usage experience (all $p > 0.05$), indicating good baseline equivalence (see Table 1).

Table 1. Descriptive statistics of baseline characteristics of participants in the two experiments.

Characteristic	Experiment 1 Human Assistance (n=106)	Experiment 1 LLM Assistance (n=107)	Experiment 2 Batch-Response LLM (n=107)	Experiment 2 Constrained-Response LLM (n=106)
Age (Mean ± SD)	22.3 ± 1.9	22.5 ± 1.7	23.0 ± 2.0	23.2 ± 2.2
Male (%)	43.4	45.8	42.1	40.6
Bachelor's Degree or Higher (%)	91.5	90.7	93.5	93.4
LLM Experience > 6 months (%)	68.9	70.1	72.0	70.8

4.2. Descriptive Statistics

Table 2 presents the descriptive statistics of the main variables for Experiment 1. The objective design quality was highest in the LLM-assisted/simple task group (M = 3.578, SD = 0.545) and lowest in the LLM-assisted/complex task group (M = 2.625, SD = 0.771).Cognitive activation was highest in the LLM-assisted/simple task group (M = 3.891, SD = 0.583), whereas design dependency was highest in the LLM-assisted/complex task group (M = 1.894, SD = 0.607).

Table 2. Descriptive Statistics of Main Variables in Experiment 1.

Condition	n	Design Quality (Mean ± SD)	Cognitive Activation (Mean ± SD)	Design Dependency (Mean ± SD)	Self-Rated Design Quality (Mean ± SD)
Human Assistance — Simple Task	53	3.428 ± 0.659	3.121 ± 0.614	1.408 ± 0.433	3.755 ± 0.691
LLM Assistance — Simple Task	54	3.578 ± 0.545	3.891 ± 0.583	1.463 ± 0.559	3.248 ± 0.807

Human Assistance — Complex Task	53	3.037 ± 0.722	2.871 ± 0.661	1.515 ± 0.565	3.577 ± 0.651
LLM Assistance — Complex Task	53	2.625 ± 0.771	2.541 ± 0.631	1.894 ± 0.607	3.270 ± 0.695

In Experiment 2, the descriptive statistics of the main variables for each condition are presented in Table 3. The objective design quality in the constrained-response LLM/complex task group (M = 3.042, SD = 0.723) was significantly higher than that in the batch-response LLM/complex task group (M = 2.568, SD = 0.679).

Table 3. Descriptive Statistics of Main Variables in Experiment 2.

Condition	n	Design Quality (Mean ± SD)	Cognitive Activation (Mean ± SD)	Design Dependency (Mean ± SD)	Condition
Batch-Response — Simple Task	53	3.671 ± 0.632	3.937 ± 0.551	1.532 ± 0.517	Batch-Response — Simple Task
Constrained-Response — Simple Task	53	3.452 ± 0.569	3.413 ± 0.675	1.355 ± 0.489	Constrained-Response — Simple Task
Batch-Response — Complex Task	54	2.568 ± 0.679	2.493 ± 0.617	1.853 ± 0.545	Batch-Response — Complex Task
Constrained-Response — Complex Task	53	3.042 ± 0.723	2.671 ± 0.732	1.474 ± 0.501	Constrained-Response — Complex Task

4.3. Data Preprocessing

The raw data were preprocessed using the following steps:

- Outlier detection: Extreme values were identified using the three standard deviations criterion. A total of seven potential outliers were detected; after individual inspection, all were deemed within the normal range of variation and retained;
- Missing value handling: Three cases with missing design quality scores due to equipment failure were removed using listwise deletion; no other variables had missing data；
- Normality test: The Shapiro—Wilk test was applied to the main variables under each condition, indicating that all variables satisfied the assumption of normality (all $p > 0.05$), meeting the requirements for parametric analyses；
- Homogeneity of variance: Levene's test indicated no significant differences in variance between groups (all $p > 0.05$), satisfying the assumption of homogeneity of variance for ANOVA.

5. Results

5.1. Experiment 1 Results

Manipulation Check: Independent-samples t-tests confirmed the effectiveness of the task complexity manipulation. Participants rated the perceived complexity significantly higher in the complex task condition compared to the simple task condition (M_complex = 3.40, SD = 0.90 vs. M_simple = 2.57, SD = 0.82; $t_{(202)} = 6.12$, $p < 0.001$, $d = 0.86$, 95% CI: [0.56, 1.10]), indicating a successful manipulation of task complexity.

Self-Rated Design Quality: A one-way ANOVA revealed a significant main effect of assistance type on self-rated design quality ($F_{1,202} = 14.37$, $p < 0.001$, $\eta^2\_p = 0.066$). Participants in the human-assisted group rated their own design performance significantly higher than those in the LLM-assisted group (M_human = 3.666, SD = 0.672 vs. M_LLM = 3.259, SD = 0.751; $t_{(202)} = 3.79$, $p < 0.001$, $d = 0.531$, 95% CI: [0.196, 0.618]). Further analysis showed that the difference between self-rated and objectively scored design quality was significantly larger in the human-assisted group than in the LLM-assisted group (M_human = 0.574, SD = 0.712 vs. M_LLM = 0.448, SD = 0.681; $t_{(202)} = 4.97$, $p < 0.001$, $d = 0.697$), indicating a stronger self-overestimation tendency in the human-assisted group (see Figure 2).

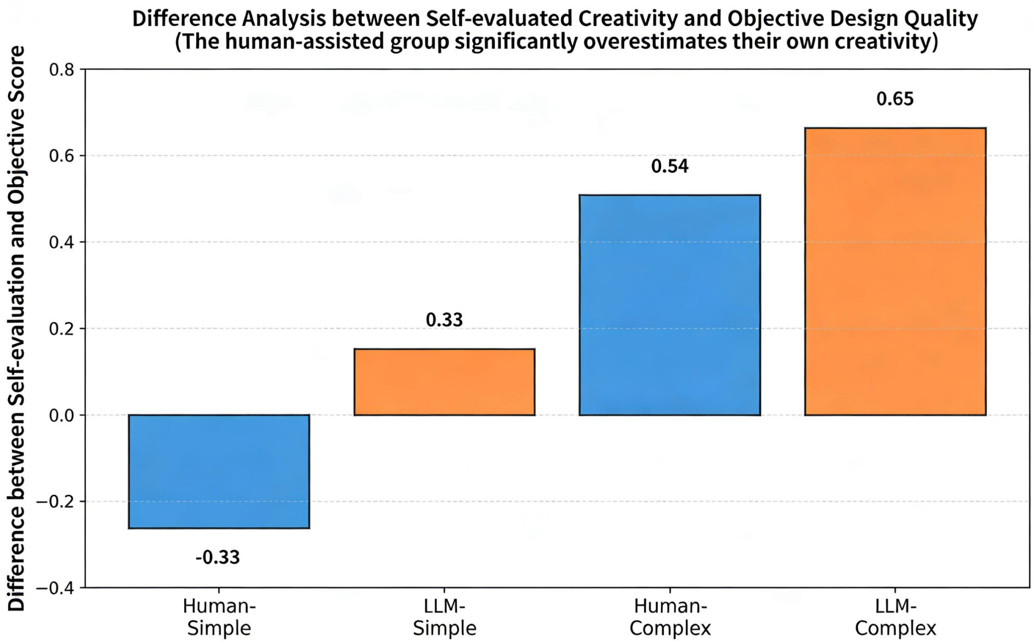

Figure 2. Analysis of the difference between self-rated and objective design quality across conditions, showing a more pronounced self-overestimation tendency in the human-assisted group.

Objective Design Quality — Interaction Effects: A 2 × 2 ANOVA revealed a significant main effect of assistance type ($F_{1,200} = 4.90$, $p = 0.028$, $\eta^2\_p = 0.024$) and a significant interaction between assistance type and task complexity ($F_{1,200} = 13.27$,

$p < 0.001$, $\eta^2\_p = 0.062$), whereas the main effect of task complexity was not significant ($F_{1,200} = 0.23$, $p = 0.631$, $\eta^2\_p = 0.001$). As shown in Figure 3, for simple design tasks, the objective design quality in the LLM-assisted group was significantly higher than that in the human-assisted group (M_LLM = 3.578, SD = 0.545 vs. M_human = 3.428, SD = 0.659; $F_{1,200} = 17.15$, $p < 0.001$, $\eta^2\_p = 0.079$). In contrast, for complex design tasks, the LLM-assisted group exhibited significantly lower objective design quality than the human-assisted group (M_LLM = 2.625, SD = 0.771 vs. M_human = 3.037, SD = 0.722; $F_{1,200} = 1.02$, $p = 0.028$, $\eta^2\_p = 0.005$). These results provide support for H1 and H2.

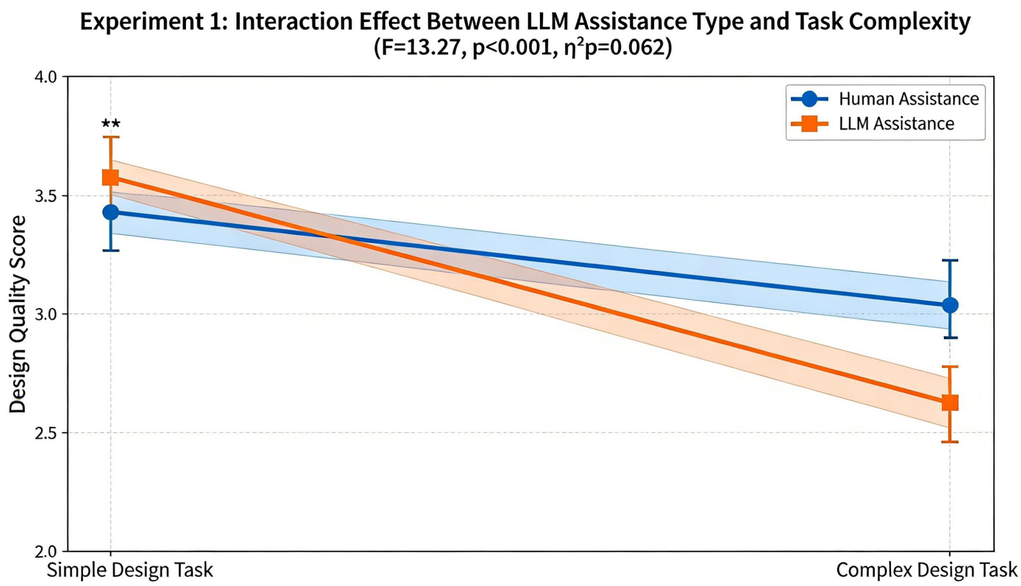

Figure 3. Interaction effects of LLM assistance type and task complexity on objective design quality in Experiment 1. Error bars represent standard errors. ** $p < 0.01$, * $p < 0.05$.

Distribution of Cognitive Activation: As shown in Figure 4, the LLM-assisted/simple task group exhibited the highest level of cognitive activation, with a median of 3.89 and an interquartile range (IQR) of 0.82, whereas the LLM-assisted/complex task group showed the lowest level, with a median of 2.54 and an IQR of 0.87. No significant outliers were observed in any group, indicating that the data distributions were robust..

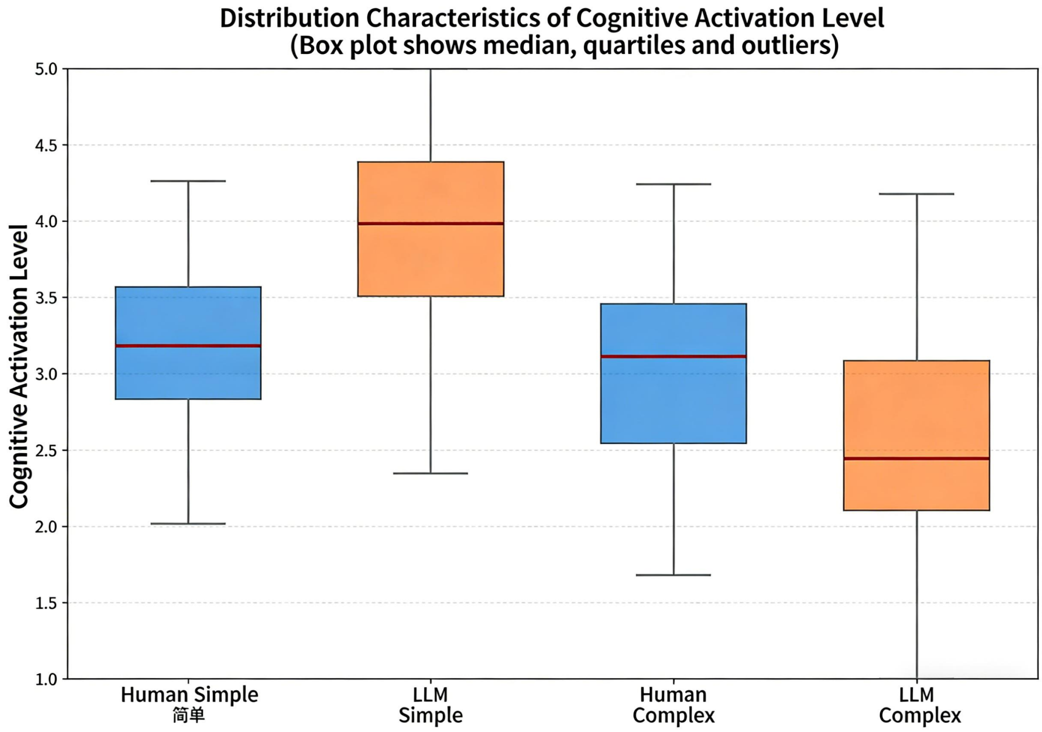

Figure 4. Boxplots of cognitive activation levels across the four experimental conditions, showing medians, quartiles, and data variability.

Distribution of Design Dependency: As shown in Figure 5, the LLM–assisted/complex task group exhibited the highest level of design dependency, with a distribution showing pronounced positive skewness, indicating that a subset of participants in this condition demonstrated extremely high design dependency. In contrast, the human–assisted groups across both task complexity levels exhibited relatively lower design dependency, with more concentrated distributions.

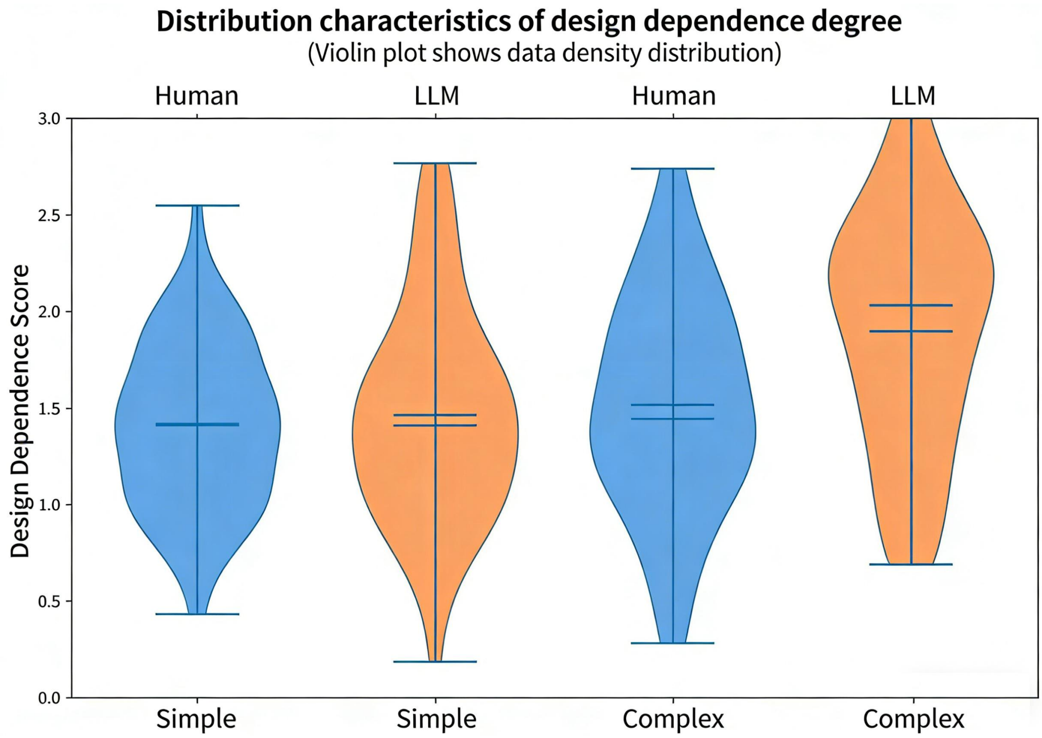

Figure 5. Violin plots depicting the distribution of design dependency across the four experimental conditions, illustrating the density and spread of the data.

Moderated Mediation Analysis: Moderated mediation analysis was conducted using the PROCESS macro (Model 8), with 10,000 bootstrap resamples to estimate confidence intervals for the indirect effects (see Figure 6). The results indicated that under the complex task condition, LLM assistance (relative to human assistance) significantly increased design dependency ($\beta = 1.449$, SE = 0.154, $p < 0.001$), and design dependency was significantly negatively associated with design quality ($\beta = -0.222$, SE = 0.072, $p = 0.002$). The conditional indirect effect of design dependency was significantly negative in the complex task condition ($\omega M1 = -0.322$, 95% CI: [−0.536, −0.120]), but not significant in the simple task condition ($\omega M1 = -0.003$, 95% CI: [−0.044, 0.045]). The index of moderated mediation was significant ($\omega M1 = -0.319$, 95% CI: [−0.551, −0.118]), supporting H2 and H3.

Under the simple task condition, LLM assistance significantly enhanced cognitive activation ($\beta = 0.891$, SE = 0.147, $p < 0.001$), and cognitive activation was significantly positively associated with design quality ($\beta = 0.341$, SE = 0.068, $p < 0.001$). The conditional indirect effect of cognitive activation was significantly positive in the simple task condition ($\omega M2 = 0.304$, 95% CI: [0.152, 0.478]), but not significant in the complex task condition ($\omega M2 = 0.012$, 95% CI: [−0.038, 0.067]). The index of moderated mediation was significant ($\omega M2 = 0.292$, 95% CI: [0.108, 0.492]), supporting H1 and H3.

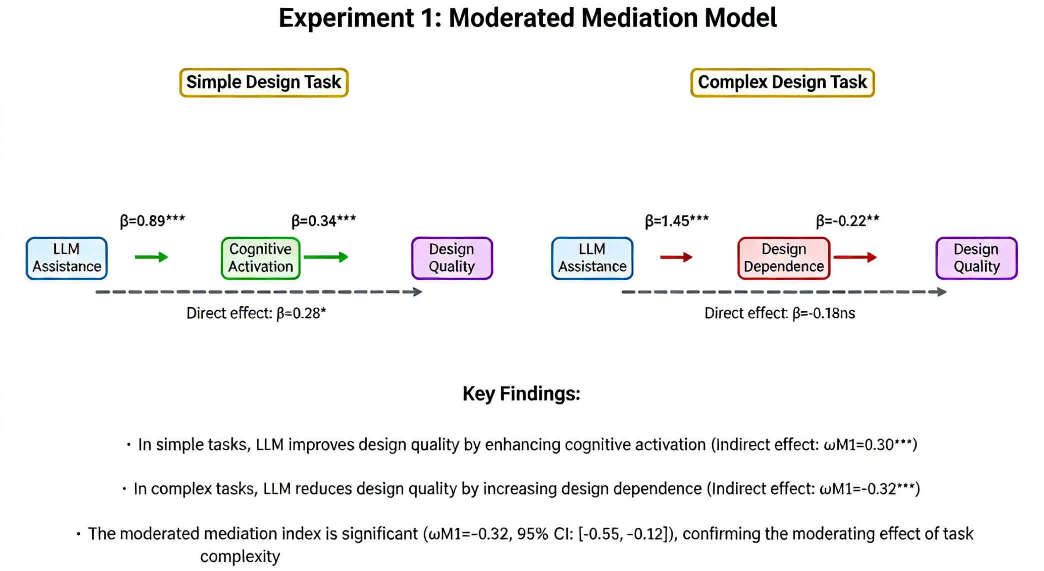

Figure 6. Boxplots of cognitive activation levels across the four experimental conditions, showing medians, quartiles, and data variability.

Sensitivity Analysis: All analyses were repeated after excluding the three cases with missing data, and the results were highly consistent with the primary analyses (the directions and significance levels of all effects remained unchanged), indicating that the main conclusions are robust. The correlations among the variables are presented in Figure 7.

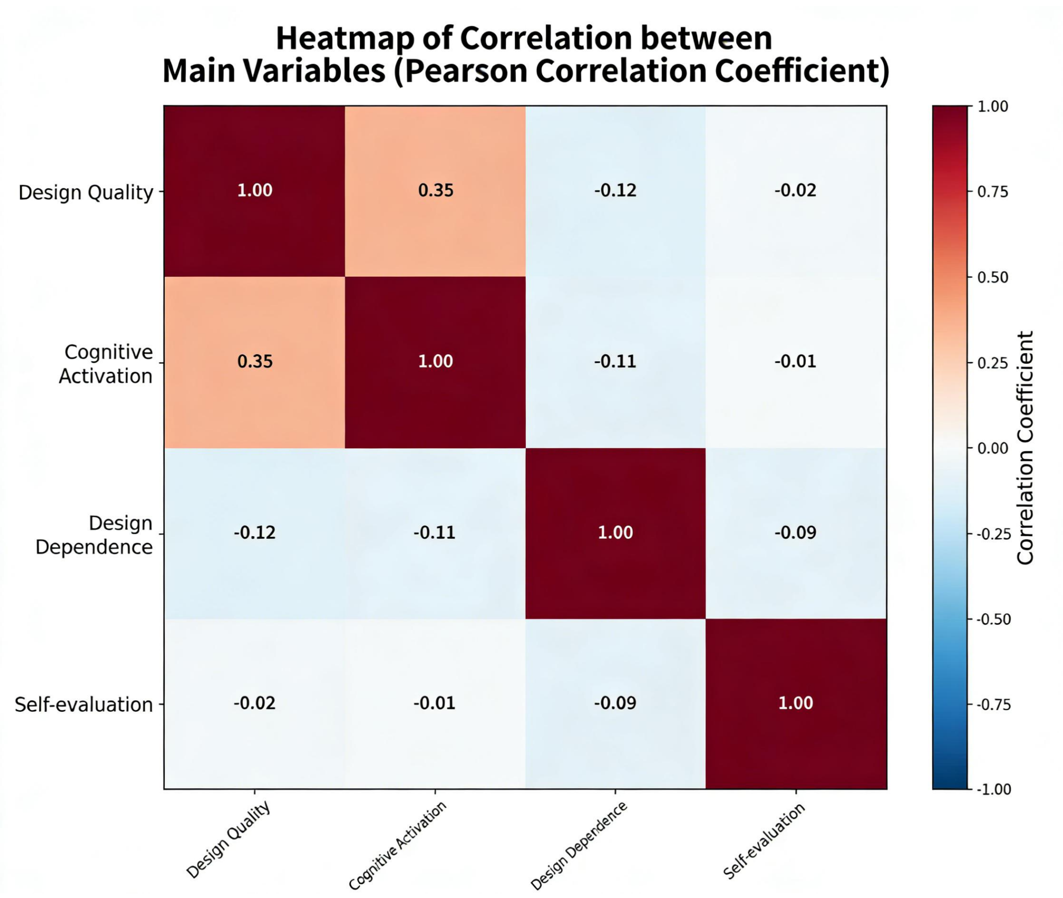

Figure 7. Pearson correlation heatmap of the main variables, illustrating the relationships among design quality, cognitive activation, design dependency, and self-rated design quality.

5.2. Experiment 2 Results

Manipulation Check: The manipulation check for task complexity was consistent with Experiment 1. Participants rated the perceived complexity significantly higher in the complex task condition than in the simple task condition (M_complex = 3.38, SD = 0.91 vs. M_simple = 2.61, SD = 0.83; $t_{(202)} = 5.87$, $p < 0.001$, d = 0.82, 95% CI: [0.51, 1.03]).

Interaction Effects on Objective Design Quality: A 2 × 2 ANOVA revealed a significant interaction between LLM response type and task complexity ($F_{1,200} = 18.54$, $p < 0.001$, $\eta^2\_p = 0.081$). As shown in Figure 6, for complex design tasks, the constrained-response LLM group exhibited significantly higher design quality than the batch-response LLM group (M_constrained = 3.042, SD = 0.723 vs. M_batch = 2.568, SD = 0.679; $F_{1,200} = 16.23$, $p < 0.001$, $\eta^2\_p = 0.075$). Conversely, for simple design tasks, the constrained-response LLM group showed significantly lower design quality than the batch-response LLM group (M_constrained = 3.452, SD = 0.569 vs. M_batch = 3.671, SD = 0.632; $F_{1,200} = 4.87$, $p = 0.028$, $\eta^2\_p = 0.024$).

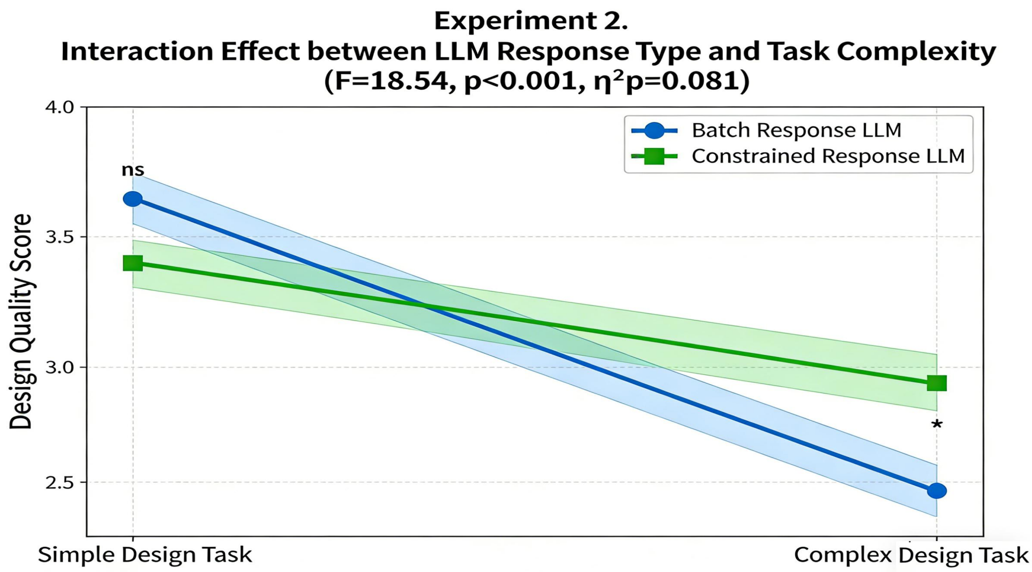

Figure 8. Interaction effects of LLM response type and task complexity on design quality in Experiment 2. Error bars represent standard errors. * p < 0.05; ns = not significant.

Moderated Mediation Analysis: As shown in Figure 8, under the complex task condition, constrained-response LLMs (compared to batch-response LLMs) significantly reduced design dependency ($\beta = -1.433$, SE = 0.159, $p < 0.001$), and design dependency was significantly negatively associated with design quality ($\beta = -0.259$, SE = 0.071, $p < 0.001$). The conditional indirect effect of design dependency was significantly positive in the complex task condition (ωM1 = 0.371, 95% CI: [0.158, 0.624]), but not significant in the simple task condition (ωM1 = 0.018, 95% CI: [−0.043, 0.090]). The index of moderated mediation was significant (ωM1 = 0.319, 95% CI: [0.082, 0.502]), supporting H4.

Under the simple task condition, constrained-response LLMs significantly reduced cognitive activation (β = −0.907, SE = 0.182, $p < 0.001$), and cognitive activation was significantly positively associated with design quality (β = 0.256, SE = 0.068, $p < 0.001$). The conditional indirect effect of cognitive activation was significantly negative in the simple task condition (ωM2 = −0.233, 95% CI: [−0.458, −0.084]), but not significant in the complex task condition (ωM2 = −0.012, 95% CI: [−0.067, 0.103]). The index of moderated mediation was significant (ωM2 = 0.244, 95% CI: [0.075, 0.502]). These results provide support for H4.

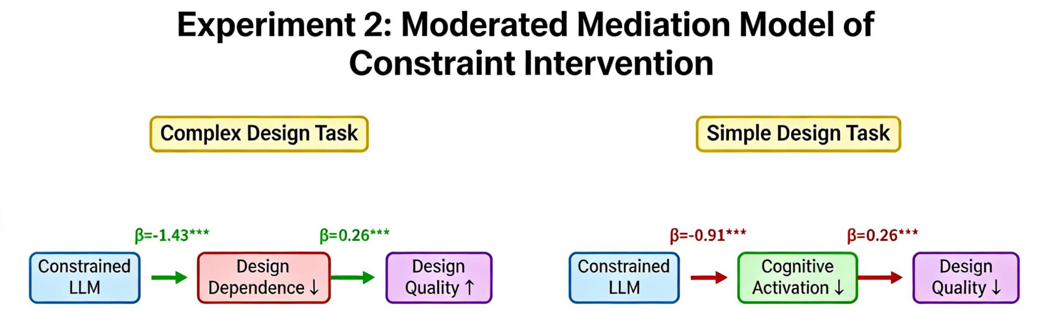

Figure 9. Moderated mediation model of the constrained-response intervention in Experiment 2, illustrating the differential effects of the constrained strategy under different task complexity conditions.

Distribution of Design Quality: Figure 10 presents histograms of design quality scores across the four conditions in Experiment 1. The distributions in all conditions were approximately normal. The LLM-assisted/simple task group showed the highest peak (mean = 3.578), while the LLM-assisted/complex task group exhibited the greatest dispersion (SD = 0.771), reflecting increased individual differences under the complex task condition.

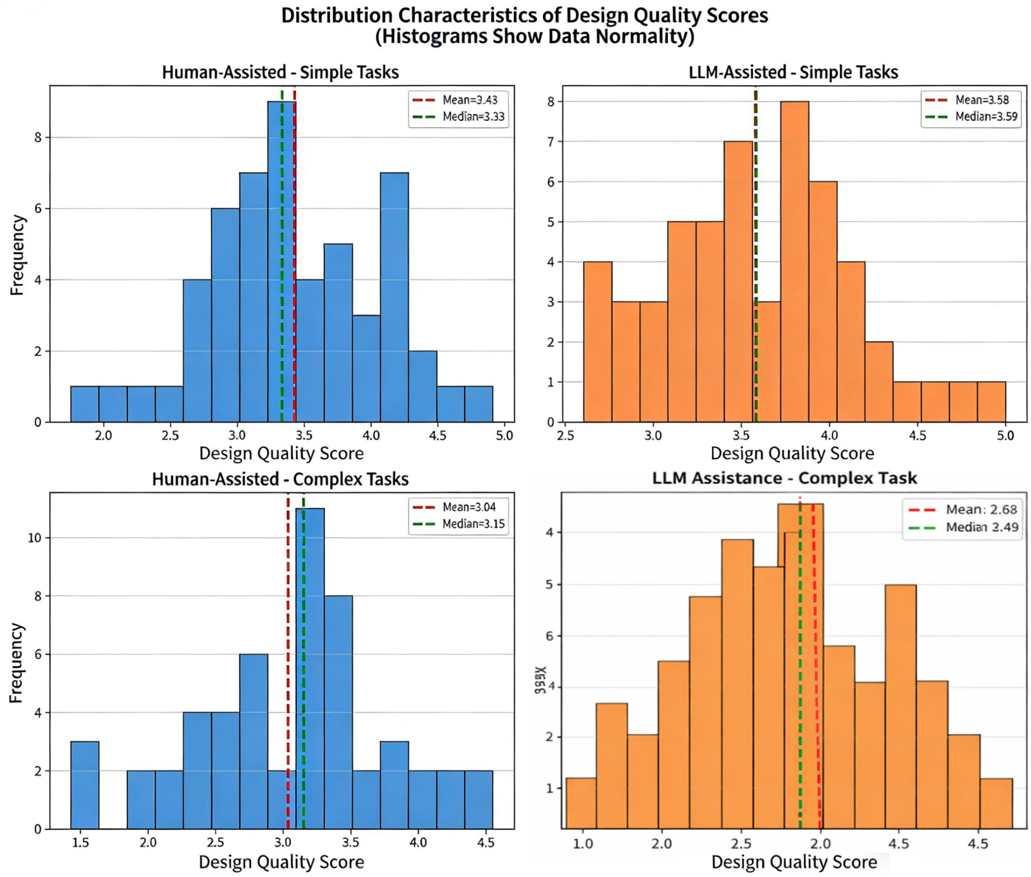

Figure 10. Histograms of design quality scores across the four conditions in Experiment 1. Red dashed lines indicate the mean, and green dashed lines indicate the median.

Multidimensional Indicator Radar Chart: Figure 11 presents a radar chart illustrating the comparative performance of the four experimental conditions across four dimensions: design quality, cognitive activation, self-rated design quality, and design dependency (reversed). The LLM-assisted/simple task condition demonstrated the highest performance in design quality and cognitive activation, whereas the LLM-assisted/complex task condition performed worst on the reversed design dependency dimension, consistent with the statistical analysis results.

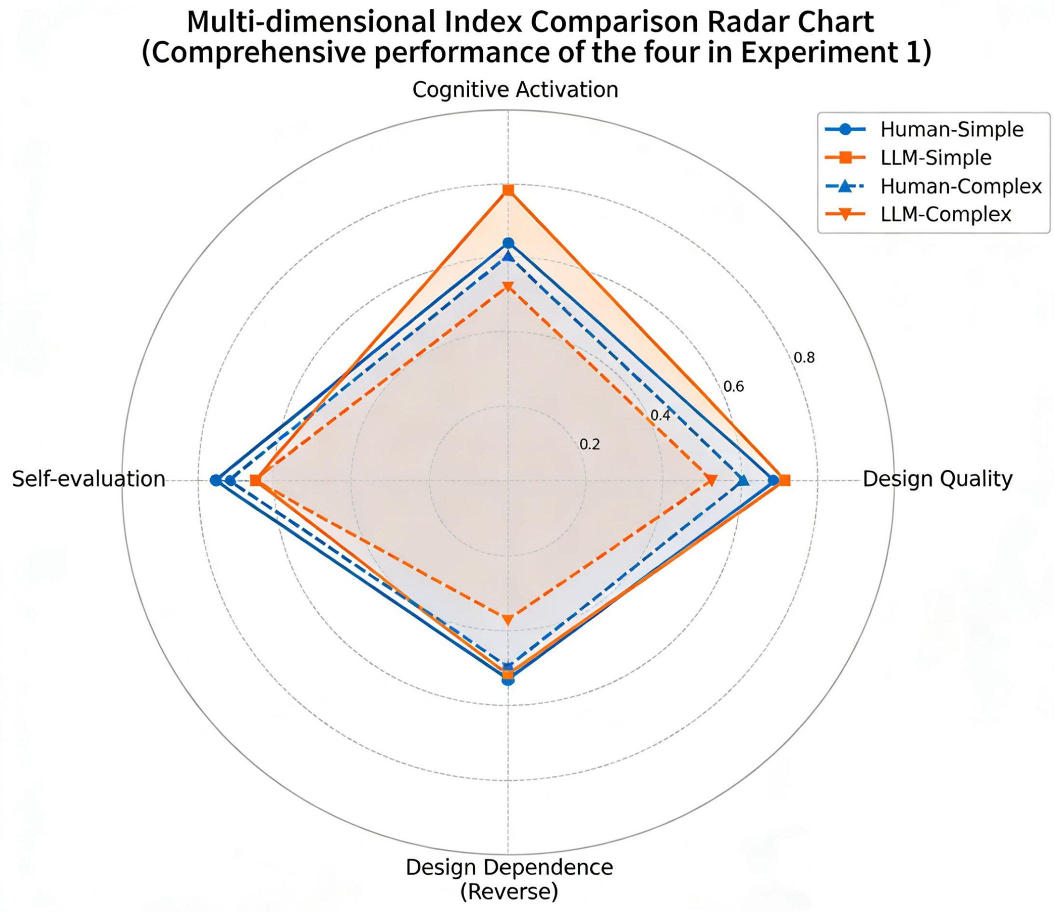

Figure 11. Radar chart comparing the four experimental conditions across multiple dimensions, highlighting the relative strengths and weaknesses of each condition.

Comparison of Effect Sizes: Figure 12 summarizes the effect sizes of the main statistical effects across the two experiments, providing a visual comparison of their relative strengths. The interaction effect of response type × task complexity in Experiment 2 (η^2_p = 0.081) was slightly larger than the interaction effect of assistance type × task complexity in Experiment 1 (η^2_p = 0.062), indicating that the constrained–response intervention has considerable practical significance.

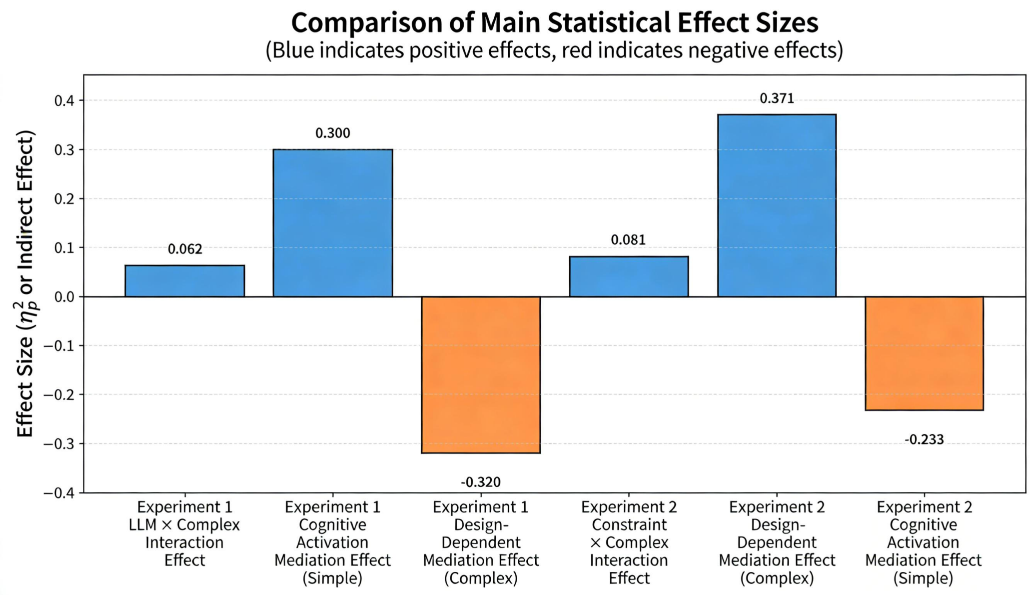

Figure 12. Bar chart comparing the effect sizes of the main statistical effects across the two experiments. Blue bars represent positive effects, and red bars represent negative effects.

6. Discussion

6.1. Dual-Opposing Mechanisms of LLMs on Design Quality

The core findings of this study reveal a dual-opposing mechanism through which LLMs influence design quality. In simple design tasks, LLM assistance significantly enhances design quality by activating cognitive associations. In contrast, in complex design tasks, LLM assistance reduces design quality due to the induction of design dependency. This finding aligns closely with prior research by Cheng and Zhang (2025) in creative task contexts, while demonstrating a unique manifestation in the specific context of design tasks.

From the perspective of Cognitive Load Theory, the formation of this dual mechanism follows a clear theoretical logic. In simple design tasks, the intrinsic cognitive load of the task is low, leaving designers' working memory with ample residual capacity. The rich design references provided by LLMs function as a "cognitive catalyst," activating remote associations and cross-domain conceptual transfer, thereby helping designers overcome fixation and generate more innovative solutions. This mechanism corroborates findings by Doshi and Hauser regarding LLMs' facilitation of remote associative thinking [26], and aligns with Urban et al.'s observation that ChatGPT enhances performance in creative problem-solving tasks [25].

However, in complex design tasks, the situation changes fundamentally. The high intrinsic cognitive load of these tasks leaves designers' working memory nearly saturated. The large volume of structured outputs generated by LLMs, rather than effectively offloading intrinsic load, further occupies limited cognitive resources in the form of extraneous load, ultimately triggering cognitive overload. Under this overloaded state, designers tend to perceive the solutions provided by LLMs as “good enough,” reducing independent cognitive effort and leading to design dependency. This finding is consistent with Hofstetter et al.’s research on information overload causing creative fixation [39], providing a novel cognitive mechanism to explain design fixation phenomena.

Notably, the study also identified an interesting “self-evaluation bias”: participants in the human-assisted group rated their own design quality significantly higher than those in the LLM-assisted group, even though objective design quality did not show the same pattern. This suggests that the “superhuman performance” of LLMs—such as rapidly generating a large number of highly structured solutions—may establish an implicit “high benchmark,” thereby undermining designers’ self-efficacy and leading them to subjectively underestimate their own capabilities, despite unchanged objective performance. This mechanism is consistent with Bandura’s self-efficacy theory [40] and has important practical implications for design education.

6.2. Differential Effects of Constrained Interventions

The results of Experiment 2 further confirmed the effectiveness of constrained LLM outputs in complex design tasks. By reducing the number of suggestions per output, constrained-response LLMs effectively lowered designers’ extraneous cognitive load, thereby mitigating cognitive overload and the formation of design dependency, ultimately improving design quality. This finding aligns with Hofstetter et al.’s research indicating that constraining others’ creative outputs can reduce creative fixation [39], and extends it to the novel context of LLM-assisted design.

From the perspective of the exploration—exploitation mechanism, the effect of constrained interventions can be further theoretically explained. Design innovation inherently involves a dynamic cycle between exploration (broad, open-ended idea generation) and exploitation (deep, focused refinement) [41]. In complex tasks, batch-response LLMs providing a large number of suggestions can lead designers to excessive exploration. Constrained-response LLMs, by limiting outputs, encourage designers to shift toward exploitation, reducing ineffective exploration and alleviating design dependency. This interpretation is consistent with Tromp’s theoretical framework on the role of constraints in the creative process [41].

However, constrained interventions produced the opposite effect in simple design tasks, resulting in reduced design quality. This finding indicates that the effectiveness of constrained strategies is highly context-dependent and cannot be generalized. In simple tasks, designers’ working memory has sufficient residual capacity, and abundant external information can effectively activate cognitive associations. Constraining LLM outputs reduces exposure to external information, thereby suppressing cognitive activation and decreasing design quality. This finding has important practical implications: the configuration of design tools should be tailored according to task complexity rather than applying a “one-size-fits-all” strategy.

6.3. Comparison with Existing Research and Attribution of Differences

The findings of this study exhibit both convergence and divergence with existing research. Compared to the study by Lee and Chung (2024), this study similarly found significant effects of LLMs on design quality, but further revealed the moderating role of task complexity on the direction of these effects [28]. Lee and Chung did not differentiate task complexity, which may have caused positive and negative effects to cancel each other out, thereby underestimating the actual impact of LLMs. By introducing task complexity as a moderating variable, the present study not only uncovers a more nuanced mechanism but also provides a framework for explaining inconsistencies in prior findings.

Compared to Anderson et al.’s research on LLM-induced creative homogenization, the present study observed a similar phenomenon at the individual level, termed “design dependency,” but attributed it to cognitive overload rather than social influence mechanisms [27]. This difference may arise from contextual distinctions: Anderson et al. focused on long-term, group-level effects, whereas the present study examined individual-level effects in single interactions. These mechanisms may operate independently or reinforce each other over prolonged use, which warrants further investigation in future research.

The findings of this study partially diverge from Bouschery et al.’s conclusion that human—AI collaboration outperforms purely human collaboration [42]. While Bouschery et al. primarily investigated open-ended creative tasks, the present study found that in complex design tasks, LLM assistance was less effective than human assistance. This difference may be attributed to task type: open-ended creative tasks involve relatively low intrinsic cognitive load, whereas complex design tasks impose high intrinsic load, making the extraneous load induced by LLM outputs more prominent.

7. Conclusion

7.1. Core Findings

- This study, based on two controlled experiments, systematically revealed a dual-opposing mechanism through which LLMs influence individual design quality in design tasks. The findings demonstrate that the impact of LLMs on design quality is not unidirectional, but rather arises from two opposing pathways: cognitive activation (positive pathway) and design dependency (negative pathway). The net effect of LLM assistance is determined by task complexity. In simple design tasks, the cognitive activation effect predominates, and LLM assistance enhances design quality. In complex design tasks, the design dependency effect dominates, and LLM assistance reduces design quality. Furthermore, constraining LLM outputs effectively mitigates design dependency in complex tasks but simultaneously weakens cognitive activation in simple tasks, indicating that the effectiveness of constraint strategies is task-specific.

7.2. Theoretical Contributions and Practical Implications

Theoretically, this study makes three primary contributions:

- It extends Cognitive Load Theory to the context of LLM-assisted design tasks and proposes a dual-opposing mechanism model of LLM effects on design quality, providing a novel framework for understanding human—AI collaborative design processes;
- It systematically examines, for the first time, the moderating role of task complexity on LLM-assisted design outcomes, addressing a gap in existing research;
- It introduces and validates the concept of "design dependency" within the design domain, enriching the theoretical framework of design fixation research.

Practically, the findings offer actionable guidance for design educators and practitioners:

- For simple design tasks, designers should be encouraged to leverage the rich references provided by batch-response LLMs to maximize the cognitive activation effect;
- For complex design tasks, constrained-response LLMs or phased introduction of LLM assistance should be adopted to prevent cognitive overload and the formation of design dependency;
- In design education, students should be trained to develop a critical understanding of LLM assistance and cultivate metacognitive skills to flexibly adjust LLM usage strategies according to task complexity.

7.3. Limitations

This study has several limitations. First, regarding scope, the participants were primarily students with basic design knowledge, whose LLM experience and professional design skills were relatively limited. Therefore, the generalizability of the findings to professional designers requires further validation. Second, regarding methodology, the study employed a single LLM system (based on the GPT-4 architecture). Differences across LLMs in terms of output diversity, structure, and other characteristics may affect the generalizability of the results. Third, regarding data, the duration of the experimental tasks was limited to 10 minutes, which may not fully capture the dynamic processes of human—AI collaboration in real-world design practice. Fourth, the study did not account for individual differences among designers (e.g., cognitive style, design experience) that may moderate the effects of LLM assistance, which warrants investigation in future research.

7.4. Future Research Directions

Based on the above limitations and findings, several directions for future research are proposed. First, investigate the differential effects of LLM assistance on designers with varying levels of expertise (novices vs. experts) to clarify the moderating role of professional design knowledge. Second, conduct longitudinal studies to track the long-term impact of LLM assistance on designers' independent design capabilities and to assess the cumulative effects of design dependency. Third, explore dynamic constraint strategies, whereby the level of LLM output constraint is adjusted according to the design task stage (conceptual exploration vs. solution refinement) to optimize the balance between cognitive activation and design dependency. Fourth, extend the research context to other design domains, such as industrial design, architectural design, and service design, to examine the cross-domain generalizability of the dual-opposing mechanism.

References

[1] Lee BC, Chung J (2024) An empirical investigation of the impact of ChatGPT on creativity. Nature Human Behaviour, 8:1906—1914. https://doi.org/10.1038/s41562-024-01953-1

[2] Anderson BR, Shah JH, Kreminski M (2024) Homogenization effects of large language models on human creative ideation. In: Proceedings of the 16th Conference on Creativity and Cognition. ACM, pp 413—425.

[3] Eapen T, Finkenstadt DJ, Folk J, Venkataswamy L (2023) How generative AI can augment human creativity. Harvard Business Review, 101:55—64.

[4] Boussioux L, Lane JN, Zhang M, Jacimovic V, Lakhani KR (2024) The crowdless future? Generative AI and creative problem solving. Organization Science, 35:1589—1607.

[5] Urban M, Děchtěrenko F, Lukavský J, Hrabalová V, Svacha F, Brom C (2024) ChatGPT improves creative problem-solving performance in university students: an experimental study. Computers & Education, 215:105031.

[6] Rafner J, Beaty RE, Kaufman JC et al. (2023) Creativity in the age of generative AI. Nature Human Behaviour, 7:1836—1838.

[7] Doshi AR, Hauser OP (2024) Generative AI enhances individual creativity but reduces the collective diversity of novel content. Science Advances, 10:eadn5290.

[8] Cross N (2011) Design Thinking: Understanding How Designers Think and Work. Berg Publishers, Oxford.

[9] Jia N, Luo X, Fang Z, Liao C (2024) When and how artificial intelligence augments employee creativity. Academy of Management Journal, 67:5—32.

[10] Liu Q, Zhou Y, Huang J, Li G (2024) When ChatGPT is gone: creativity reverts and homogeneity persists. Preprint at https://doi.org/10.48550/arXiv.2401.06816

[11] Nakadai R, Nakawake Y, Shibasaki S (2023) AI language tools risk scientific diversity and innovation. Nature Human Behaviour, 7:1804—1805.

[12] Crilly N (2019) Creativity and fixation in the real world: a literature review of case study research. Design Studies, 64:154—168.

[13] Gallupe RB et al. (1992) Electronic brainstorming and group size. Academy of Management Journal, 35:350—369.

[14] Janssen J, Kirschner PA (2020) Applying collaborative cognitive load theory to computer-supported collaborative learning: towards a research agenda. Educational Technology Research and Development, 68:783—805.

[15] Kirschner PA, Sweller J, Kirschner F, Zambrano RJ (2018) From cognitive load theory to collaborative cognitive load theory. International Journal of Computer-Supported Collaborative Learning, 13:213—233.

[16] Sweller J (2010) Element interactivity and intrinsic, extraneous, and germane cognitive load. Educational Psychology Review, 22:123—138.

[17] Kirschner F, Paas F, Kirschner PA (2011) Task complexity as a driver for collaborative learning efficiency: the collective working-memory effect. Applied Cognitive Psychology, 25:615—624.

[18] Goldschmidt G (2014) Linkography: Unfolding the Design Process. MIT Press, Cambridge.

[19] Jiang H, Yen CC (2009) Cognitive load in design problem solving. In: Proceedings of the International Conference on Engineering Design, pp 1—12.

[20] Cross N (2004) Expertise in design: an overview. Design Studies, 25:427—441.

[21] Jansson DG, Smith SM (1991) Design fixation. Design Studies, 12:3—11.

[22] Kirschner PA, Sweller J, Kirschner F, Zambrano RJ (2018) From cognitive load theory to collaborative cognitive load theory. International Journal of Computer-Supported Collaborative Learning, 13:213—233.

[23] Eapen T, Finkenstadt DJ, Folk J, Venkataswamy L (2023) How generative AI can augment human creativity. Harvard Business Review, 101:55—64.

[24] Bouschery SG, Blazevic V, Piller FT (2024) Artificial intelligence-augmented brainstorming: how humans and AI beat humans alone. Available via https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4724068

[25] Urban M et al. (2024) ChatGPT improves creative problem-solving performance in university students: an experimental study. Computers & Education, 215:105031.

[26] Doshi AR, Hauser OP (2024) Generative AI enhances individual creativity but reduces the collective diversity of novel content. Science Advances, 10:eadn5290.

[27] Anderson BR, Shah JH, Kreminski M (2024) Homogenization effects of large language models on human creative ideation. In: Proceedings of the 16th Conference on Creativity and Cognition. ACM, pp 413—425.

[28] Lee BC, Chung J (2024) An empirical investigation of the impact of ChatGPT on creativity. Nature Human Behaviour, 8:1906—1914.

[29] Rafner J, Beaty RE, Kaufman JC et al. (2023) Creativity in the age of generative AI. Nature Human Behaviour, 7:1836—1838.

[30] Wood DJ (1983) Studies in the Quality of Life: Thinking and Feeling. Falmer Press, London.

[31] Kirschner F, Paas F, Kirschner PA (2011) Task complexity as a driver for collaborative learning efficiency. Applied Cognitive Psychology, 25:615—624.

[32] Stempfle J, Badke-Schaub P (2002) Thinking in design teams: an analysis of team communication. Design Studies, 23:473—496.

[33] Elen J, Clark RE (2006) Handling Complexity in Learning Environments: Theory and Research. Emerald Group Publishing.

[34] Paas F, Renkl A, Sweller J (2003) Cognitive load theory and instructional design: recent developments. Educational Psychologist, 38:1—4.

[35] Salimzadeh S, He G, Gadiraju U (2024) Dealing with uncertainty: understanding the impact of prognostic versus diagnostic tasks on trust and reliance in human-AI decision making. In: Proceedings of the CHI Conference on Human Factors in Computing Systems. ACM, pp 1—17.

[36] Böttger T, Rudolph T, Evanschitzky H, Pfrang T (2017) Customer inspiration: conceptualization, scale development, and validation. Journal of Marketing, 81:116—131.

[37] Lu JG, Akinola M, Mason MF (2017) "Switching On" creativity: task switching can increase creativity by reducing cognitive fixation. Organizational Behavior and Human Decision Processes, 139:63—75.

[38] Hayes AF (2015) An index and test of linear moderated mediation. Multivariate Behavioral Research, 50:1—22.

[39] Hofstetter R, Dahl DW, Aryobsei S, Herrmann A (2021) Constraining ideas: how seeing ideas of others harms creativity in open innovation. Journal of Marketing Research, 58:95—114.

[40] Bandura A (1997) Self-Efficacy: The Exercise of Control. W.H. Freeman, New York.

[41] Tromp C (2024) Creativity from constraint exploration and exploitation. Psychological Reports, 127:1818—1843.

[42] Bouschery SG, Blazevic V, Piller FT (2024) Artificial intelligence-augmented brainstorming: how humans and AI beat humans alone. SSRN Working Paper.