applied sciences

MDPI

Article

From Manual Ideation to AI-Augmented Exploration: Evaluating Human–AI Workflows in Furniture Design Education (A Quasi-Experimental Study)

Dana Khalid Amro

Faculty of Architecture and Design, Al Ahliyya Amman University, Amman P.O. Box 19111, Jordan; d.amro@ammanu.edu.jo

Abstract

Generative artificial intelligence has emerged as a tool that can influence students' ideation, development, and communication of design solutions. A quasi-experimental mixed-methods study investigated the application of three generative AI tools, specifically Vizcom, Krea.ai 1, and ChatGPT 4, to furniture design in an undergraduate interior design course offered at a private Jordanian university. This study sought to address a gap in region-specific AI tools for design education. Thirty-four third-year students completed both a manual design activity (pre-test) and a redesign activity with AI assistance (post-test) after structured training. Twenty-seven participants completed both activities and were included in the analyses. Experience was measured with the Creativity Support Index (CSI) and NASA Task Load Index (NASA-TLX); qualitative measures included open-ended questions and expert-juror ratings of the 27 resulting designs. Quantitative results were analyzed with descriptive statistics and Wilcoxon signed-rank tests (SPSS v28). AI-assisted workflows significantly enhanced exploration ($Z = -3.42$, $p < 0.001$, $r = 0.66$) and approached significance for workflow efficiency ($Z = -1.97$, $p = 0.049$, $r = 0.38$). Students reported decreases in mental, time, and effort burden, while feelings of expressiveness and ownership remained. Jury experts concluded that 67.6% of students achieved improved outcomes using AI tools. Students experienced moderate frustration with prompt creation, tools operating in English, and AI's inability to design intricate details. Generative AI can be most beneficial during the conceptual stages, enabling broader exploration and greater efficiency. With careful pedagogy and ethical discussion, students can use AI tools without losing ownership of their designs. Consideration for generative AI software implementation in interior design programs in Jordan and across the MENA region is provided.

Keywords: generative AI (GenAI); interior design education; furniture design; Creativity Support Index (CSI); NASA Task Load Index (NASA-TLX); cognitive workload; creative process; AI in education

Check for updates

Academic Editor: Andrea Prati

Received: 3 May 2026
Revised: 4 June 2026
Accepted: 11 June 2026
Published: 23 June 2026

Copyright: © 2026 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Generative artificial intelligence (generative AI) technologies are reshaping how we think about creativity and design instruction by changing how students conceptualize, visualize, and produce design solutions [1,2]. Interior Design is among the most experimental and iterative majors taught within design schools. This is due to the significant volume of spatial reasoning, material intelligence, aesthetic translations, and client relations students need to absorb [3]. Furniture designers may sketch by hand, construct physical models to

work through ideas/concepts, and constantly test different materials/construction methods. These steps are mentally and physically taxing and time-consuming, which contributes to the cognitive load students experience [4]. Generative AI technologies like ChatGPT 4, Vizcom, and Krea.ai 1 can be leveraged to enhance ideation, enable quicker iteration, and explore ideas previously inaccessible to undergraduates [5].

As technology changes the approaches to interior design education, the potential impacts of AI design tools on design thinking, creativity, and learning outcomes should be explored and assessed critically and empirically [2,6]. Existing studies show promise while also highlighting limitations of AI-generated creative work, including student learning outcomes, prompt-related technical issues, and authorship and academic integrity [7]. Nonetheless, much of the empirical research on AI integration into design education has been conducted in either Western European or East Asian contexts. This leaves a gap in localized research on the integration of AI tools into undergraduate design education in Arab-world contexts and, more generally, across the MENA region. This paper aims to fill this gap by presenting findings from a Jordanian context that has been understudied and faces distinct cultural, linguistic, and resource constraints that may affect students' acceptance of and experience with AI technologies. The importance of coordinated human–AI interaction and multi-agent decision support, as discussed in recent Sustainability research [8], further reinforces the growing interdisciplinary relevance of AI-supported collaborative systems in design and creative problem-solving contexts. Although originating from a different disciplinary domain, the study reinforces the broader relevance of coordinated multi-agent AI systems in supporting complex human decision-making environments. Although interior design and furniture design represent distinct yet interconnected domains, the present study focuses on furniture design pedagogy within an undergraduate interior design curriculum, in which furniture studios serve as specialized sub-disciplinary design modules. This interdisciplinary perspective aligns with recent findings on collaborative AI-supported decision systems and coordination mechanisms in complex environments [8].

This research addresses a notable gap in the literature by providing empirical insight into the impacts of three generative AI tools not only on juror-rated design quality of student outputs but also on students' subjective experiences of the design process, quantitatively captured through established psychometric measures. We employ the NASA Task Load Index (NASA-TLX) [4] to quantify students' perceived cognitive workload and the Creativity Support Index (CSI) [9] to measure students' perceptions of how well each generative AI tool supported specific aspects of the creative design process. Coupled with thematic analysis of open-ended responses and expert evaluations of student design solutions, we present a nuanced methodological evaluation of AI-enhanced learning in furniture design education.

The research questions guiding this study are threefold: (RQ1) How does traditional manual design influence students' perceived cognitive workload and creativity support during furniture design briefs? (RQ2) How do generative AI tools (Vizcom, Krea.ai, ChatGPT 4) influence students' cognitive workload and creativity support? (RQ3) What qualitative impacts does generative AI integration have on output quality and students' perceptions of authorship, collaboration, and pedagogy?

2. Literature Review

2.1. Generative AI in Creative Design: Mechanisms and Capabilities

Generative AI software refers to AI programs capable of generating new content/output, such as images, text, and 3D models. They generate new data/content by learning statistical patterns from large amounts of training data. Today, generative AI

tools have become powerful design tools for designers/architects working across design disciplines such as architecture, interiors, graphic design, and industrial design. Currently, there are two main streams of generative AI technologies/design used for design purposes: text-to-image (Txt2img) generation and generative design (GD) algorithms.

Text-to-image generation is the process of producing imagery from human-written text, commonly called 'prompts'. These images are meant to convey the semantic meaning of the text. Midjourney, DALL-E, Vizcom, Krea.ai, and other leading platforms for text-to-image generation use multimodal generative models that combine diffusion architectures and large language models to produce text-to-image results. The quality and detail of the outputs depend heavily on prompt engineering: crafting accurate, precise, and contextually relevant prompts that effectively constrain the generative model's output space [10]. Crafting prompts successfully requires domain knowledge to determine which settings to use, as well as metalinguistic knowledge of the model itself [11]. Additional recent studies on AI-driven interior generation and workflow automation further support these capabilities [12–16].

Generative Design (GD) is another complementary approach in which algorithms, given a set of constraints, automatically solve technical design problems and produce families of optimized forms [17]. GD software like Midjourny, allows designers to query parametric relationships between material properties, structural behaviors, manufacturing constraints, and/or sustainability targets. Most GD software packages use unsupervised machine learning techniques, such as Generative Adversarial Networks (GANs), to infer complex structural relationships from training datasets. GD approaches are often used for constraint-led optimization; however, designers must understand the engineering and design context to provide relevant parameters and assess whether the outputs meet functional and/or aesthetic requirements [18].

2.2. Applications of Generative AI in Interior and Furniture Design

2.2.1. Interior Design Applications

Interior Design Practice also uses AI. Interior design practice applications of AI include space layout optimization, color palette design, material surveys, and interior style customization. One study explored the use of large language models and image generators for selecting and texturing materials in 3D interior spaces. They found that AI tools could provide users with relevant information about material properties (such as durability, affordability, eco-friendliness, or style match) faster than manual research via material libraries [19,20]. AI technologies were paired with Virtual Reality (VR) systems to allow users to explore AI-generated interior design proposals in immersive spatial environments interactively [21]. Another study points out that AI processing technologies can be used to automate mundane tasks in the creation of design documents while allowing designers to maintain final judgment [8,22]. Recent diffusion-model approaches and GAN-based layout systems also demonstrate the growing role of AI in automated interior generation and furniture layout optimization [12,15,16].

2.2.2. Furniture Design and AIGC Technology

AI-Generated Content (AIGC) has significant potential in the furniture industry to quickly generate concepts, improve designs, and enable personalization [6]. The specific applications of AIGC in furniture could include generative design (structural design), text-to-image creation (idea design), and AI analysis (including user preferences and marketing trend analysis). An example of a commercially released piece of furniture created with AI technologies is designer Philippe Starck's 'AI Chair', developed using generative design algorithms in collaboration with materialization partner Kartell and Autodesk. This chair

was shown to validate AI design by creating stable furniture forms while also achieving an aesthetically novel outcome by blending algorithmically generated design options with human designer intent [23]. These AI design programs can also rapidly produce many design iterations to reach visual concepts in a fraction of the time it would take designers working exclusively by hand [2].

2.3. Human–AI Collaboration: Dynamics, Benefits, and Limitations

2.3.1. The Collaborative Paradigm and Creative Benefits

Generative AI changes the designer's role from practitioner-as-author to practitioner-as-director. The generative designer collaborates with artificial intelligence design systems, working as the lead creative, rather than the AI replacing designers. The designer will always remain an important part of the process because human intuition, industry knowledge, and decision-making are irreplaceable. It is ideal for automating tedious tasks and for creating design options for humans to evaluate, choose, and modify [24]. AI tools have proven particularly effective during the ideation and early creative phases, especially for problem identification, divergent idea generation, and initial concept evaluation, yet comparatively less effective in later stages that require contextual understanding, cross-functional teamwork, and practical implementation skills [2].

It has been demonstrated that students can also benefit greatly from this collaborative model. Generative AI design tools allow students to explore additional paths when conceptualizing ideas more quickly, iterate faster, and explore ideas they might not have been able to produce as meticulously by hand [6]. Studies using the Creativity Support Index (CSI) to evaluate generative AI tools have consistently found high scores in exploration and enjoyment, reaffirming that students can broaden their range of ideas and feel more engaged with the task at hand [25–28].

2.3.2. Limitations and Challenges

Still, several significant obstacles remain to realizing these potentials in design practice. Existing text-to-image programs have shown flaws in their ability to convey spatial logic, factual accuracy, and specificity in architectural and design vocabularies, particularly when representing culturally/historically determined design languages [10]. Bias is present in AI as it is trained on data that can overrepresent hegemonic design influences and underrepresent vernacular design approaches. Outputs are highly dependent on the accuracy of the prompt, but developing successful prompts is a nuanced metalinguistic skill that requires continual refinement through experience and familiarity with the subject [10]. Similar worries have been raised about over-dependence on AI in the context of design education. Overreliance on automatically generated ideas could lead to the "atrophy" of students' fundamental design skills, such as hand sketching, spatial reasoning, and material awareness, particularly when AI tools are used independently of pedagogical practices that foreground design as an embodied activity [6]. Equity concerns also arise in the context of AI in design education. The financial and time barriers to accessing AI tools, as well as linguistic barriers for tools with English-dominated interfaces, place students in lower-resource educational contexts (and specifically Jordanian students designing in Arabic contexts) at a relative disadvantage.

2.4. Generative AI in Design Education: Curricular Integration and Pedagogical Implications

2.4.1. Curricular Integration and the Pedagogical Shift

In response to AI's emergence into professional design practice, there is a need to expand design curricula [1,6]. Studio courses that replicate the professional practice of interior design need to include relevant tools to prepare students for practice [1]. Advocates

for integrating AI into design curricula suggest that designers treat AI as they have past technologies, such as CAD, photorealistic rendering, and BIM. Technical skills can be slowly developed throughout their educational programs [1]. In an example of a curriculum-embedded case [2], they introduced AI tools into an existing furniture design course curriculum that centered on material and construction techniques. Students implemented Midjourney and DALL-E tools to create conceptual designs. By controlling its introduction, authors were able to expose students to the generative possibilities and misinterpretation problems of AI tools.

2.4.2. Creativity Enhancement, Skill Development, and Authorship

AI-based tools have been implemented as Creativity Support Tools (CSTs) in university-level design curricula. These systems facilitate both divergent thinking (idea generation) and convergent thinking (idea evaluation and development into feasible solutions). Research has found that students who used AI-based visual programming tools created more diverse and novel designs than those working without AI assistance [5]. Ownership and authorship over creative work are major considerations when implementing AI into the design education process. Research indicates that learners often do not view AI applications as a replacement for human creativity [1]. Learners will often view whatever an AI program produces as a starting point they can review and edit as needed, leaning on their design intuition to take what they need from the AI and transform it into their projects [1]. Schools can encourage students to maintain academic integrity while demystifying AI as a creative tool by normalizing good practices such as asking students to mark when they use AI tools, to avoid copying and pasting prompts directly, and to clearly identify which parts of their design process were AI-generated [2,29].

2.4.3. Ethical and Equity Considerations

Ethical issues can also arise when using generative AI. Aside from worries about academic integrity when using generative AI tools for schoolwork, there are other concerns, such as data privacy, algorithmic bias, and equal access to technology. AIGC requires substantial high-quality training data to work. Commercial AIGC platforms often use proprietary datasets that reflect certain biases. When these generators are used to create design work that falls outside of Western style and taste, the outputs can be systematically biased [6]. Teachers, therefore, need to guide students in questioning AI outputs, spotting bias, evaluating cultural relevance, and resisting the temptation to treat them as unbiased or objective. Institutions also need to consider issues of copyright and ownership with AI tools. While some AI platforms are more transparent about the sources of their training data than others, students and educators have varying rights to the content they create with these tools [2].

2.5. Measuring Creativity and Cognitive Workload in AI-Assisted Design

2.5.1. NASA Task Load Index

The NASA Task Load Index (NASA-TLX) is a reliable, multidimensional subjective workload assessment tool that evaluates the perceived workload of tasks participants are asked to perform [4,30]. Originally, NASA-TLX comprised six subscales, including mental demand, physical demand, temporal demand, performance, effort, and frustration. The subscales are assessed using weighted pairwise comparisons and subscale ratings, yielding an overall task load score ranging from 0 to 100. When performing tasks that involve little physical activity (such as computer-based design work), the physical demand subscale can be ignored. NASA-TLX has been successfully applied to measuring workload during AI-assisted design tasks, such as digital design tool evaluation [5,31,32].

2.5.2. Creativity Support Index

Creativity Support Index (CSI) is a validated psychometric tool used to quantify support for salient components of creative practice [9]. CSI yields both a factor score (max 20) indicating how well supported a creative dimension is by the tool and a factor count (max 5) indicating the importance of that dimension for a given creative endeavor. The six dimensions evaluated by the instrument include exploration, expressiveness, results worth the effort (efficiency), collaboration, immersion, and enjoyment. Results from CSI deployments in design education settings and AI-tool evaluation experiments have shown that exploration and enjoyment receive the highest ratings when working with AI tools, which reaffirms that participants explore more ideas and experience positive emotions [25,33]. Table 1 provides an overview of notable studies that have used the CSI in design contexts [34].

Table 1. Representative studies employing the NASA-TLX and CSI in design and AI-assisted Creative contexts.

Reference	Framework	Dimensions Assessed	Key Findings
Remy et al. (2020) [12]	CSI	All six dimensions	Benchmarked creativity tools in HCI design contexts; exploration and enjoyment are consistently the top priorities.
Chandrasekera et al. (2024) [5]	CSI + NASA-TLX	All CSI + five NASA-TLX subscales	GenAI enhances creativity during early design stages; workload decreases; exploration shows the greatest improvement.
Paananen et al. (2024) [15]	CSI	Exploration, Expressiveness	AI-assisted ideation in architecture: enhanced exploration, expressiveness and authorship preserved

Note: CSI = Creativity Support Index; NASA-TLX = NASA Task Load Index. Only studies with verified publications are included.

2.6. Research Gap and Contribution of the Study

Grounded in the previous literature discussed above, our investigation contributes to the current literature on generative AI's effects on designers' creativity and cognitive load when adopted in learning environments. Two major exceptions exist in the literature. The majority of existing empirical studies focus on Western European, North American, or East Asian institutional settings. These settings shed minimal light on how generative AI adoption operates in Arab or MENA contexts, which have unique languages, cultures, and resources [1,2]. Second, only a limited number of studies have investigated the psychometrics of the student experience using validated measures (NASA-TLX, CSI) alongside the quality of design outputs (juried expert validation) within the same intervention [5]. Therefore, the current study aims to fill this gap by employing a quasi-experimental mixed-methods design within an undergraduate interior design program in Jordan, triangulating psychometric student experience data with expert evaluations of student design work to establish localized empirical evidence on the use of AI in undergraduate design education.

3. Methodology

3.1. Research Design

The current study took place in the context of furniture design education. A quasi-experimental mixed-methods design measured students' cognitive load, perceived support for creativity, and the quality of design outcomes when using generative AI design tools

versus when not. A quasi-experimental design was selected because participants could not be randomly assigned to conditions, given the nature of a single intact university course cohort [35]. Qualitative and quantitative methods are used to provide a holistic view of the students' experience. Manual versus AI-assisted workflows were directly compared through two validated psychometric assessment tools: The NASA Task Load Index (NASA-TLX) [4] and the Creativity Support Index (CSI) [9]. Additional insights were drawn from juror ratings of design outputs and from a thematic analysis of qualitative responses. Importantly, the study should be interpreted as evaluating an AI-supported iterative redesign workflow rather than a fully isolated comparison between manual and AI-generated creativity because the AI-assisted phase intentionally built upon previously developed manual concepts.

3.2. Participants

The participants were 34 third-year students enrolled in a furniture design course within their interior design program at a private university in Jordan. The students were asked to participate in the second semester of the 2024/2025 academic year. Twenty-seven participants were female (79.4%), and seven were male (20.6%), which corresponded to the typical demographic distribution in interior design programs at the university. Students reported using digital design tools at varying levels, yet none had formal training with generative AI platforms before the study. The 34 students completed the pre-test measures, and 27 completed both the pre-test and post-test surveys. Seven students were unable to complete the post-survey due to either absence or incomplete submission. Our final paired sample included n = 27. There were no significant differences in demographics or pre-survey measures between students who completed the study and those who did not (all $p > 0.10$), suggesting that attrition was not biased.

3.3. Research Procedure

The study lasted four weeks and was divided into three phases, as shown in Figure 1:

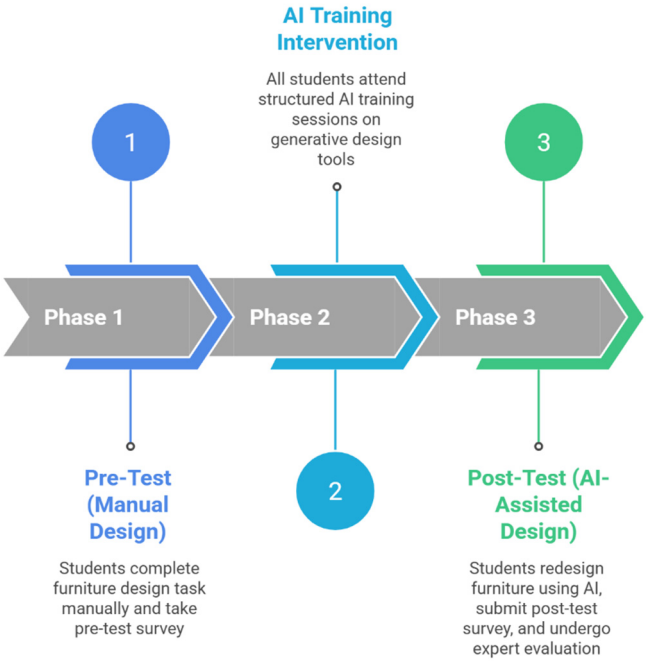


Figure 1. The research process workflow.

Phase 1—Pre-Test (Manual Design): Participants completed a furniture design project individually, using traditional methods without access to AI. Submissions included mood

boards, hand sketches, rendered design concepts, and explanations of the rationale and material selection. Students turned in their project and completed a formal pre-test consisting of survey questions that aligned with CSI dimensions and questions focused on their design process.

Phase 2—AI Training Intervention: Students received instruction on using AI tools in two subsequent sessions after completing the pre-test. A professional AI tutor facilitated both sessions. Session one introduced students to generative AI design tools. The training involved live demonstrations of Vizcom, Krea.ai, and ChatGPT 4. Session two consisted of individual feedback, follow-up, and practice with prompt engineering. All students received the same level of intervention and support using these three AI platforms.

Phase 3—Post-Test (AI-Enhanced Design): In this phase, students revised their furniture designs using the AI tool. Students created AI-generated concept imagery, updated digital prototypes, and finalized material-selection documentation as part of their post-test submissions. In total, 27 participants completed the post-test questionnaire, which was identical to the pre-test survey except for questions about participants' experiences with and perceptions of the AI. An independent juror with design expertise reviewed all student submissions (n = 27 drawings total from both phases).

In summary, the conceptual framework (Figure 2) of this study was based on an iterative human–AI furniture design workflow integrating manual conceptual development, AI-assisted exploration, reflective evaluation, and final design resolution. Initially, students developed furniture concepts manually through sketching, functional thinking, and ergonomic considerations. Subsequently, generative AI tools, including Vizcom, Krea.ai, and ChatGPT 4, were used to support rapid visualization, exploratory ideation, iterative variation generation, and rendering enhancement. Students then critically evaluated and refined AI-generated outputs through human judgment and design reasoning, maintaining authorship and decision-making throughout the process. Final furniture proposals were assessed using the CSI, NASA-TLX, and expert juror evaluation. The framework, therefore, positions generative AI as an exploratory co-design and augmentation system rather than a replacement for human creativity in furniture design education.



Figure 2. The conceptual framework of this study.

3.4. Research Instruments

3.4.1. Quantitative Instruments

Likert items (1 = Strongly Disagree to 5 = Strongly Agree) were included in both surveys based on NASA-TLX and CSI dimensions. There were 14 items on each survey. The pre-test survey did not contain NASA-TLX subscale questions. As such, pre-test workload was measured indirectly through open-ended comments and Likert-scale items assessing their experience with the manual design, rather than via the NASA-TLX procedure. This represents a methodological limitation, discussed further in Section 3.7, and indicates that pre-test and post-test NASA-TLX values should be regarded as comparative indicators rather than as directly paired measurements. To minimize inflation risk in interpreting workload reduction, pre-test values were treated only as indicative baseline trends and were not included in formal inferential NASA-TLX statistical comparisons. Similarly, pre-test CSI items measured anticipated perceptions of AI-supported creativity rather than experienced creativity support, meaning that RQ1 primarily establishes baseline expectations and atti-

tudes before AI exposure rather than equivalent experiential creativity outcomes. To avoid systematically inflating the apparent reduction in cognitive demand, pre-test workload indicators were interpreted descriptively only and were excluded from inferential NASA-TLX comparisons. Consequently, RQ1 should be interpreted as measuring students' baseline expectations and attitudes toward AI-supported creativity before exposure rather than directly comparing equivalent creative experiences between conditions. This distinction may partially confound shifts in attitude toward AI with genuine changes in experienced creativity support, and future longitudinal studies should administer fully equivalent CSI measures at both time points. Therefore, CSI findings during the AI-assisted phase should be interpreted primarily as indicators of exploratory support, iterative refinement, and visualization enhancement, rather than as fully independent concept generation. This interpretation aligns with human–AI co-creation frameworks that position generative AI as an augmentation system that supports reflective iteration and the expansion of ideation, rather than as an autonomous source of original creativity.

3.4.2. Qualitative Instruments

Qualitative data were collected using three complementary methods. First, open-ended survey questions encouraged students to describe their experiences with the design process, their perceived challenges, their satisfaction with the outcomes, and their reflections on their creative and decision-making practices. Second, informal interviews were held with a subset of students during the design process to explore emerging issues in real time. Third, a systematic content analysis was performed on all submitted student design outputs (n = 27 individual drawings spanning both phases), which were independently evaluated by two expert jurors with professional experience in furniture design and design education, and classified using a three-category impact rubric.

3.5. Conceptual Framework and Operationalization

NASA Task Load Index (NASA-TLX): Five subscales were assessed during the post-test phase (mental demand, temporal demand, performance, effort, and frustration [4,30]), with physical demand omitted since the task involved only seated, computer-based activity.

Creativity Support Index (CSI): Six subdimensions were assessed during both phases: creativity (the generation of new ideas), exploration (the extent of concept investigation), efficiency (including speed of iteration and ease of use), expressiveness (a sense of design ownership and creative voice), enjoyment (the emotional engagement with the design process), and collaboration (peer interaction enabled or hindered by the tool) [9].

3.6. Statistical Analysis

Quantitative data were analyzed using SPSS Version 28 [36]. Descriptive statistics, including means and standard deviations, were calculated for all pre- and post-test survey items. For the inferential comparison of pre- and post-test CSI scores across the 27 paired participants, Wilcoxon signed-rank tests were used as the primary statistical test because Shapiro–Wilk tests indicated non-normal distributions in several Likert-scale items (all items, n = 27, $W < 0.93$). Note: Effect sizes for Wilcoxon tests are presented as $r = Z/\sqrt{N}$. All tests were two-tailed, with $p < 0.05$. Before inferential testing, Shapiro–Wilk normality tests were conducted for all paired CSI variables. Results indicated non-normal distributions across several variables (W values ranged from 0.81 to 0.93; $p < 0.05$), supporting the use of the nonparametric Wilcoxon signed-rank test rather than the parametric paired-samples *t*-test.

Qualitative data from open-ended responses and interview notes underwent inductive thematic analysis following Braun and Clarke's [37] six-phase process: data familiarization, initial code creation, theme development, theme review, theme definition, and reporting.

Themes were independently coded by both researchers, with intercoder agreement evaluated using Cohen's kappa (κ = 0.81, indicating strong agreement) [38]. Content analysis of design outputs was performed by two expert jurors, who independently classified each drawing into one of three categories: (1) AI significantly improved the design compared to the manual version; (2) AI produced results similar to the manual design; or (3) AI did not improve or negatively affected the manual design. Inter-rater agreement for the juror classification was Cohen's κ = 0.76.

3.7. Limitations

Several limitations of the present study should be noted: The sample consisted of 34 students from a single institution in Jordan, limiting the generalizability of the findings to other educational settings, cultures, or types of institutions. Although the seven students who dropped out between the pre- and post-test phases (20.6%) do not appear to have introduced bias, this reduction in sample size affects the statistical power of the Wilcoxon tests and paired analysis. The pre-test NASA-TLX indicators were collected indirectly through qualitative and Likert-scale responses rather than being directly administered, so the comparison of pre- and post-test NASA-TLX, should be considered suggestive rather than conclusive. Future studies should administer NASA-TLX directly at both time points. The four-week timeframe may not capture long-term learning retention or improvements in design ability. Lastly, this research focused on three AI platforms (Vizcom, Krea.ai 1, and ChatGPT 4) available during data collection; these findings may not generalize to other programs. Because the pre-test NASA-TLX values were inferred indirectly rather than directly administered, measurement bias may have been introduced through subjective interpretation of qualitative responses and Likert-scale indicators. Consequently, pre- and post-test workload comparisons should be interpreted with caution as indicative trends rather than as fully equivalent psychometric measurements.

Additionally, because the AI-assisted phase followed the manual design phase using the same furniture project, the study is susceptible to sequence and incubation effects. Improvements observed during the AI-assisted phase may reflect iterative design maturation, increased familiarity with the project, and prolonged reflection, rather than the isolated effect of generative AI tools. Accordingly, the findings should be interpreted as evaluations of AI-supported iterative redesign processes rather than as strictly isolated comparisons of AI versus manual design performance.

4. Results and Analysis

4.1. Framework Mapping and Survey Instrument Structure

Survey questions asked during the pre-test (design task without generative AI tools) and post-test (design task with generative AI tools) were aligned with categories from NASA-TLX and CSI to measure any potential differences in cognitive workload and perceptions of creativity support when generative AI tools are available. See Table 2 for a list of questions paired with each dimension and an explanation of why they were paired.

Table 2. Mapping of pre-test and post-test survey items to NASA-TLX and CSI framework dimensions.

Dimension	Framework	Pre-Test Item	Post-Test Item	Analytical Purpose
Creativity	CSI	I believe AI tools enhance creativity.	Did AI enhance your creativity?	Captures AI's perceived role in novel idea generation
Exploration	CSI	I am familiar with generative AI tools.	Where did you use AI in your design process?	Evaluates openness to and actual breadth of AI tool use

Table 2. *Cont.*

Dimension	Framework	Pre-Test Item	Post-Test Item	Analytical Purpose
Expressiveness	CSI	I would still feel design ownership using AI tools.	Did you feel ownership of your design using AI?	Indicates whether AI supports or threatens individual creative control
Efficiency	CSI	I expect AI tools to improve my workflow.	Did AI streamline your workflow?	Measures time-saving and productivity dimensions
Efficiency (Iteration)	CSI	AI can enhance iteration efficiency.	Did AI help you iterate more efficiently?	Assesses iteration speed and agility of design development
Collaboration	CSI	AI tools can enhance design team collaboration.	Did you use AI collaboratively? Did it help or hinder?	Assesses AI's role in peer-to-peer learning
Enjoyment	CSI	I enjoy experimenting with digital tools.	Did you enjoy using AI tools in your design process?	Evaluates affective and motivational engagement
Mental Demand	NASA-TLX	Inferred from manual process descriptions	AI reduced my mental effort in design.	Indicates required mental effort under each condition
Temporal Demand	NASA-TLX	Inferred from ideation time descriptions	AI made the process more time-efficient.	Reflects perceived time pressure or time savings
Performance	NASA-TLX	Inferred from satisfaction and ownership reports	AI improved my design effectiveness.	Measures satisfaction with design output quality
Effort	NASA-TLX	Inferred from manual task difficulty descriptions	Integrating AI was easy.	Measures perceived energy expenditure under each condition
Frustration	NASA-TLX	Inferred from open-ended frustration reports	I faced limitations using AI tools.	Reflects emotional strain or tool-related obstacles

Note: CSI = Creativity Support Index; NASA-TLX = NASA Task Load Index. NASA-TLX pre-test questions were inferred indirectly from qualitative and Likert-scale data; NASA-TLX post-test questions were administered explicitly. This represents an imbalance between conditions (as detailed in Section 3.7, Limitations).

4.2. Pre-Test Results: Manual Design Phase

All 34 participants completed the pre-test survey. Table 3 shows descriptive statistics for CSI-aligned items. To ensure consistency with the post-test presentation, all CSI descriptive statistics are reported using the same format (Mean = M; Standard Deviation = SD) across both Sections 4.2 and 4.3.

Table 3. Descriptive statistics for pre-test CSI survey items (manual design phase; n = 34).

CSI Dimension	Survey Item	M	SD
Creativity	I believe AI tools will enhance creativity.	3.97	0.83
Exploration	I am familiar with generative AI tools.	3.53	0.79
Expressiveness	I would still feel ownership of the design using AI tools.	3.79	0.91
Efficiency	I expect AI tools to improve my design workflow.	3.67	0.70
Efficiency (Iteration)	AI can enhance iteration efficiency.	3.91	0.77
Collaboration	AI tools can enhance design team collaboration.	3.79	0.88
Enjoyment	I enjoy experimenting with digital tools in design.	3.67	0.75

Note: Items measured on a 5-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree). n = 34.

Means from pre-test answers generally indicate positive attitudes towards AI creation. Creativity had the highest pre-test mean (M = 3.97, SD = 0.83), suggesting participants had high expectations for the creative generativity AI would offer. Exploration had the

lowest mean (M = 3.53, SD = 0.79), and many participants had little prior experience with generative AI websites (only 26.5% reported regularly using AI technology before the study). Students reported 'Efficiency (Iteration)' as having the second-highest potential value (M = 3.91). This value indicates students expected AI to have the greatest impact on the pace/breadth of the ideation process. 'Expressiveness' and 'Collaboration' had moderate-to-high valence scores (M = 3.79). This high rating indicates students did not anticipate losing creative control and were mildly optimistic about AI collaboration.

Table 4 shows estimated NASA-TLX scores based on qualitative evaluation of open-ended responses concerning experience with manual design before testing began. (Refer to Section 3.7 for clarification.) Please note that these are rough baseline values and are not intended to represent actual NASA-TLX scores.

Table 4. Approximate NASA-TLX Indicators inferred from pre-test open-ended responses (manual design phase; n = 34).

NASA-TLX Dimension	Qualitative Basis (Representative Patterns)	Approx. Mean (1–5)
Mental Demand	Students described sustained cognitive effort in generating initial concepts; difficulty organising thoughts without external reference tools	3.08
Temporal Demand	Majority of students reported that manual ideation required 1–3 days for initial concept generation	3.00
Effort	Students described intensive manual sketching, physical prototyping, and repeated material reference searches as demanding processes	3.26
Frustration	Open-ended responses indicated confusion, slow progress, and lack of confidence in concept development	2.89
Performance	Inferred from self-reported satisfaction and sense of ownership over completed manual designs	3.44

Note: These values are estimates calculated from Likert-scale questions and written comments, not actually from administered NASA-TLX scores. The physical demand subscale was not administered during the study. Consider it as approximate only.

Qualitative Corroboration, Pre-Test Baseline: The approximate NASA-TLX indicators in Table 4 are supported in thematic analysis of students' open-ended pre-test responses (intercoder agreement: Cohen's $\kappa = 0.81$). Student's consistently described the manual design phase as mentally demanding: representative responses referred to "difficulty organizing initial ideas without external visual references" and "spending 1 to 3 days generating a concept that felt uncertain." These patterns confirm the inferred Mental Demand ($M \approx 3.08$) and Effort ($M \approx 3.26$) scores. The Performance score ($M \approx 3.44$) reflects moderate self-reported satisfaction with manual outputs, a baseline against which post-test AI-assisted performance gains (NASA-TLX $\Delta = +0.63$) are meaningfully interpreted. Qualitative responses also revealed that students' pre-test attitudes toward AI were generally positive but untested, with 73.5% reporting no prior regular experience with generative AI platforms, making the post-test CSI improvements particularly notable as learning-driven gains rather than prior familiarity effects.

4.3. Post-Test Results: AI-Assisted Design Phase

n = 27 students responded to the post-test questionnaire after the AI-assisted design process. Table 5 displays descriptive statistics for the CSI and NASA-TLX factors as administered directly on the post-test measure. The descriptive statistics for both CSI and NASA-TLX dimensions follow a unified reporting structure using Mean (M) and Standard Deviation (SD) values to maintain methodological consistency throughout the analysis.

Table 5. Descriptive statistics for post-test survey items—CSI and NASA-TLX dimensions (AI-assisted phase; n = 27).

Dimension	Framework	Survey Item	M	SD
Exploration	CSI	AI allowed me to explore more design alternatives.	4.11	0.76
Creativity	CSI	AI enhanced my creativity.	4.15	0.72
Expressiveness	CSI	I felt a sense of ownership over the design using AI.	4.00	0.81
Efficiency	CSI	AI streamlined my design workflow.	3.96	0.80
Efficiency (Iteration)	CSI	AI helped me iterate more efficiently.	4.04	0.74
Enjoyment	CSI	I enjoyed using AI tools in my design process.	3.89	0.77
Collaboration	CSI	AI facilitated peer collaboration.	3.85	0.84
Mental Demand	NASA-TLX	AI reduced the mental effort required in design.	3.89	0.73
Temporal Demand	NASA-TLX	AI made the design process more time efficient.	4.00	0.68
Effort	NASA-TLX	Integrating AI into my workflow was manageable.	3.85	0.72
Performance	NASA-TLX	AI improved the effectiveness of my design output.	4.15	0.69
Frustration	NASA-TLX	I encountered significant limitations using AI tools.	3.63	0.91
Ethical Awareness	Contextual	I understand the ethical/legal implications of AI use.	3.74	0.79

Note: Questions are answered using a 5-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree). n = 27. The Ethical Awareness question is unique to its block and does not factor into the scoring of NASA-TLX or CSI.

The post-survey shows uniformly high scores across all CSI and NASA-TLX areas. The two categories with the highest averages were Creativity (M = 4.15) and Performance (M = 4.15). Reduced temporal demand was another area that was rated highly by participants (M = 4.00). The area users struggled with most was Frustration (M = 3.63, SD = 0.91), which indicates that issues with the tool (specifically creating prompts, dealing with an English-only tool, and unreliable AI outputs for intricate design components) were the largest negative externality for participants.

4.4. Comparative Analysis: Pre-Test Versus Post-Test

4.4.1. Wilcoxon Signed-Rank Analysis: CSI Dimensions

Wilcoxon signed-rank tests were conducted on matched pre–post pairs (n = 27) from each CSI dimension. Tests were conducted to see whether participants perceived a signifi-

cant increase in creativity support from manual to AI design. See Table 6. All Wilcoxon signed-rank analyses were conducted exclusively on the 27 matched participants who completed both pre-test and post-test measures to preserve the validity of paired comparisons. Pre-test descriptive statistics in Table 6 are displayed for contextual completeness, but all Wilcoxon signed-rank analyses were calculated exclusively from the 27 matched participant pairs.

Table 6. Wilcoxon signed-rank comparison of pre-test and post-test CSI scores (n = 27).

CSI Dimension	Pre M	Pre SD	Post M	Post SD	Δ Mean	Z	*p*	r
Exploration	3.53	0.79	4.11	0.76	+0.58	−3.42	<0.001	0.66
Creativity	3.97	0.83	4.15	0.72	+0.18	−1.20	0.231	0.23
Expressiveness	3.79	0.91	4.00	0.81	+0.21	−1.27	0.205	0.24
Efficiency	3.67	0.70	3.96	0.80	+0.29	−1.97	0.049	0.38
Efficiency (Iter.)	3.91	0.77	4.04	0.74	+0.13	−0.89	0.373	0.17
Enjoyment	3.67	0.75	3.89	0.77	+0.22	−1.51	0.131	0.29
Collaboration	3.79	0.88	3.85	0.84	+0.06	−0.37	0.713	0.07

Note: Z = Wilcoxon Z-statistic. r = effect size (computed by dividing Z by the square root of n). The n in the Pre-test column = 34. The n in the Post-test column = 27. Since the Wilcoxon tests were conducted with matched pairs, n = 27, two-tailed tests with $\alpha = 0.05$.

Exploration had the largest and only statistically significant gain (mean change 0.58 points; $Z = -3.42$, $p < 0.001$, $r = 0.66$). This effect size is considered large according to conventional criteria ($r > 0.50$) [38]. AI tools substantially increased students' ability to explore ideas. Workflow efficiency also improved notably at the $\alpha = 0.05$ level ($Z = -1.97$, $p = 0.049$, $r = 0.38$). Although creativity ($\Delta = +0.18$, $p = 0.231$), expressiveness ($\Delta = +0.21$, $p = 0.205$), enjoyment ($\Delta = +0.22$, $p = 0.131$), and efficiency-iteration ($\Delta = +0.13$, $p = 0.373$) all showed consistent positive trends, they did not reach statistical significance, implying that a larger sample could likely confirm these improvements. Cooperation did not show a significant difference ($\Delta = +0.06$, $p = 0.713$), indicating that generative AI's current mode of individualized, prompt-based interaction is not intended for cooperative learning.

Qualitative Corroboration, CSI Exploration and Efficiency Gains: The statistically significant exploration gain ($Z = -3.42$, $p < 0.001$, $r = 0.66$) is strongly corroborated by qualitative evidence from student open-ended responses and informal interviews. A dominant theme in thematic analysis ($\kappa = 0.81$) was the acceleration of the ideation cycle: the majority of students reported transitioning from a 1–3 day manual concept-generation process to an in-meeting loop in which multiple design alternatives could be explored and compared within a single work session. Representative student statements described Vizcom and Krea.ai 1 as enabling them to "see ideas they couldn't have drawn by hand" and to "keep iterating without losing time." This qualitative compression of ideation time directly substantiates the workflow efficiency gain approaching significance ($Z = -1.97$, $p = 0.049$, $r = 0.38$). The non-significant gains for creativity ($\Delta = +0.18$, $p = 0.231$) and expressiveness ($\Delta = +0.21$, $p = 0.205$) are also interpreted qualitatively: students consistently reported perceiving their creative opinion as retained throughout the AI-assisted process, treating AI-generated outputs as "a starting point" subject to their own editorial judgment rather than a replacement for their design intent. The non-significant collaboration result ($\Delta = +0.06$, $p = 0.713$) is corroborated by the absence of collaborative prompting strategies in student responses; none described sharing prompts, jointly evaluating AI outputs, or co-developing concepts, confirming that individualized human–AI interaction was the dominant mode throughout the study.

4.4.2. Indicative NASA-TLX Workload Comparison

Table 7 compares indicative pre-test NASA-TLX scores (qualitative, estimated measurement; see Section 3.7) with post-test NASA-TLX scores. No inferential analysis is provided due to measurement imbalance. Arrows represent the direction of difference.

Table 7. Comparison of NASA-TLX workload metrics: before test (manual) vs. after test (AI-powered).

Dimension	Pre-Test (Approx.)	Post-Test (Direct) M	Post-Test SD	Δ
Mental Demand	3.08	2.70	0.73	−0.38
Temporal Demand	3.00	2.50	0.68	−0.50
Effort	3.26	2.70	0.72	−0.56
Frustration	2.89	2.52	0.91	−0.37
Performance	3.44	4.07	0.69	+0.63

Note: On the demand dimensions (Mental Demand, Temporal Demand, Effort, Frustration), lower scores on the post-test indicate a lower demand, and so are reverse-scored when compared to pre-test scores. Positive scores on Performance indicate higher perceived performance. Pre-test scores are estimates. Formally paired statistical tests cannot be used. See Section 3.7.

Results were consistent with the CSI results; AI-generated designs were correlated with significantly lower perceived effort ($\Delta = -0.56$), time pressure ($\Delta = -0.50$), and mental demand ($\Delta = -0.38$). Perceived performance had the largest mean increase (+0.63), suggesting that students were much more satisfied with their work when generating AI outputs than when designing something by hand. Participants also reported slightly lower frustration ($\Delta = -0.37$), and frustration had the largest standard deviation on the post-test of all subscales (SD = 0.91), suggesting that participants did not experience frustration related to the tool.

Qualitative Corroboration, Cognitive Workload Reduction: The indicative NASA-TLX reductions reported in Table 7 are demonstrated by students' open-ended post-test responses. Students consistently framed AI tools as relieving the most cognitively demanding early-stage tasks: generating initial visual forms, searching material references, and integrating functional constraints with aesthetic targets. Representative qualitative responses described "spending less energy on getting started" and "not feeling stuck at a blank page." These themes directly corroborate the largest reductions in workload ($\Delta = -0.56$), effort ($\Delta = -0.56$), and temporal demand ($\Delta = -0.50$), and align with the conceptual framework's function attributed to generative AI in the design education literature. The high alteration in frustration (SD = 0.91) is particularly well-explained by qualitative data: student responses revealed three distinct frustration clusters: those who struggled with prompt engineering in an unfamiliar metalinguistic register, those hindered by the English-only interfaces of Vizcom and Krea.ai, and those frustrated by AI's inability to render spatially accurate or constructionally feasible furniture details. These three subgroups experienced AI-related frustration very differently, which explains why the mean reduction in frustration was modest ($\Delta = -0.37$), and the standard deviation was the highest among the subscales.

4.4.3. Expert Juror Content Analysis of Design Outputs

The 37 student design outputs generated across both data collection phases were independently rated by two trained jurors. Ratings showed substantial agreement between raters (Cohen's $\kappa = 0.76$) [38]. See Table 8 for the number of outputs per impact category. Table 8 represents the authoritative quantitative juror classification framework based on the full dataset of 37 outputs, whereas Table 9 presents only representative visual

samples selected from each category for illustrative purposes. Therefore, Table 8 should be considered the authoritative quantitative classification framework, while Table 9 serves as a representative visual demonstration layer derived from selected examples within those categories.

Table 8. Expert juror classification of student design outputs (n = 34).

Category	n	%
AI improved design quality significantly compared to manual counterpart	23	67.6%
AI produced results comparable to manual design	6	17.6%
AI did not improve or detracted from manual design quality	5	14.7%

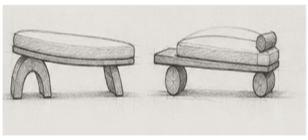
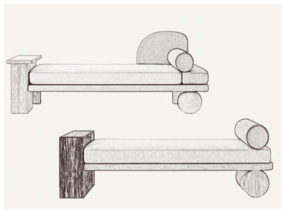
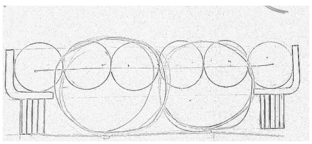
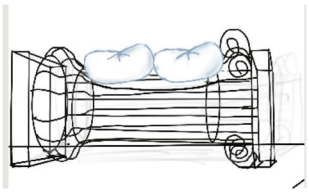
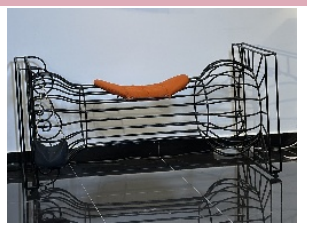
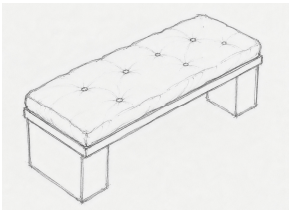
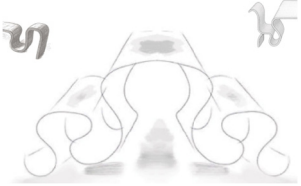
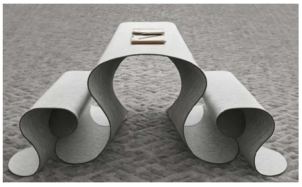
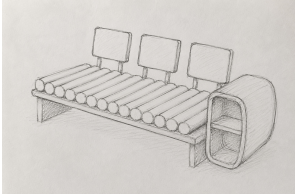
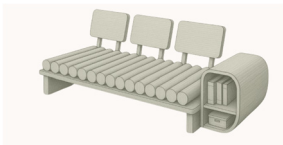
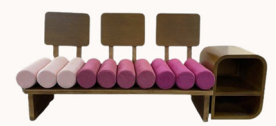
Note: Ratings were based on the agreement of two expert jurors; inter-rater reliability: Cohen's $\kappa = 0.76$ [38].

Jurors assigned to classify outcomes rated AI-produced work as superior to what the student likely would have done without AI assistance in 67.6% of cases; this was most often when there was greater variation in ideas, higher-quality graphics, or more sophisticated media. Jurors judged outcomes as about equal to what the student would likely have done without AI assistance in 17.6% of cases; this was most often when jurors felt the student's contribution was already so strong that AI would not significantly elevate the quality of the work. Jurors reported AI did not enhance design outcomes in 14.7% of cases (n = 6). Jurors explained that in these cases, students did not effectively prompt AI to generate desired outputs, were too literal in using AI outputs without customizing them, or created outputs that were too far beyond what is spatially possible or reasonable compared to manually created outputs.

Qualitative Corroboration, Expert Juror Findings: The juror categories closely align with qualitative patterns in student open-ended responses and the informal interviews conducted during the design process. In Category 1 cases (AI significantly improved, 67.6%), qualitative evidence indicates that students who accomplished the strongest outcomes engaged in active, iterative dialogue with the AI, generating multiple prompt variations, critically evaluating options, and selectively integrating outputs into their design development. Informal interviews with these students described a sense of "trying things quickly and choosing what feels right," reflecting the expanded exploration confirmed by the CSI ($\Delta = +0.58$, $p < 0.001$). In Category 2 cases (comparable results, 17.6%), qualitative responses revealed that students expressed strong prior design confidence and described AI tools as "not adding much that I hadn't already thought of", corroborating the top limit-effect interpretation. In Category 3 cases (AI detracted, 14.7%), qualitative data were most revealing: students described difficulty articulating their design intent in English-language prompts; frustration when AI outputs ignored cultural or spatial constraints specific to design practice; and, in several cases, a leaning toward accept spatially incredible AI outputs because of limited confidence in critically evaluating AI's technical output. These qualitative accounts confirm that passive AI adoption, rather than critical co-authorship, is the primary attempt condition for AI-assisted design degradation.

Figure 3 summarizes the expert juror evaluations of the AI-assisted furniture design workflow. The results show that 67.6% of students achieved significantly improved outcomes through AI-supported exploration and visualization, while 17.6% achieved results comparable to their manual designs. In contrast, 14.7% showed weaker outcomes due to ineffective prompting, passive use of AI, and limited critical evaluation. The findings highlight that successful AI integration depends on active human judgment and iterative design engagement.

Table 9. Samples of students' work outcome.

#	Design Sketch	AI Trails First Iteration	Final Results Second Iteration	Final Product
Category 1: AI Improved Design Significantly—Student Work (67.6%, n = 23)				
1				
2				
3				
Category 2: AI did not improve or detract Designer Sketch Better than AI Output (14.7%, n = 5)				
4				
5				
Category 3: AI produced comparable results Designer Sketch and AI Design Produced Similar Results (17.6%, n = 6)				
6				
7				

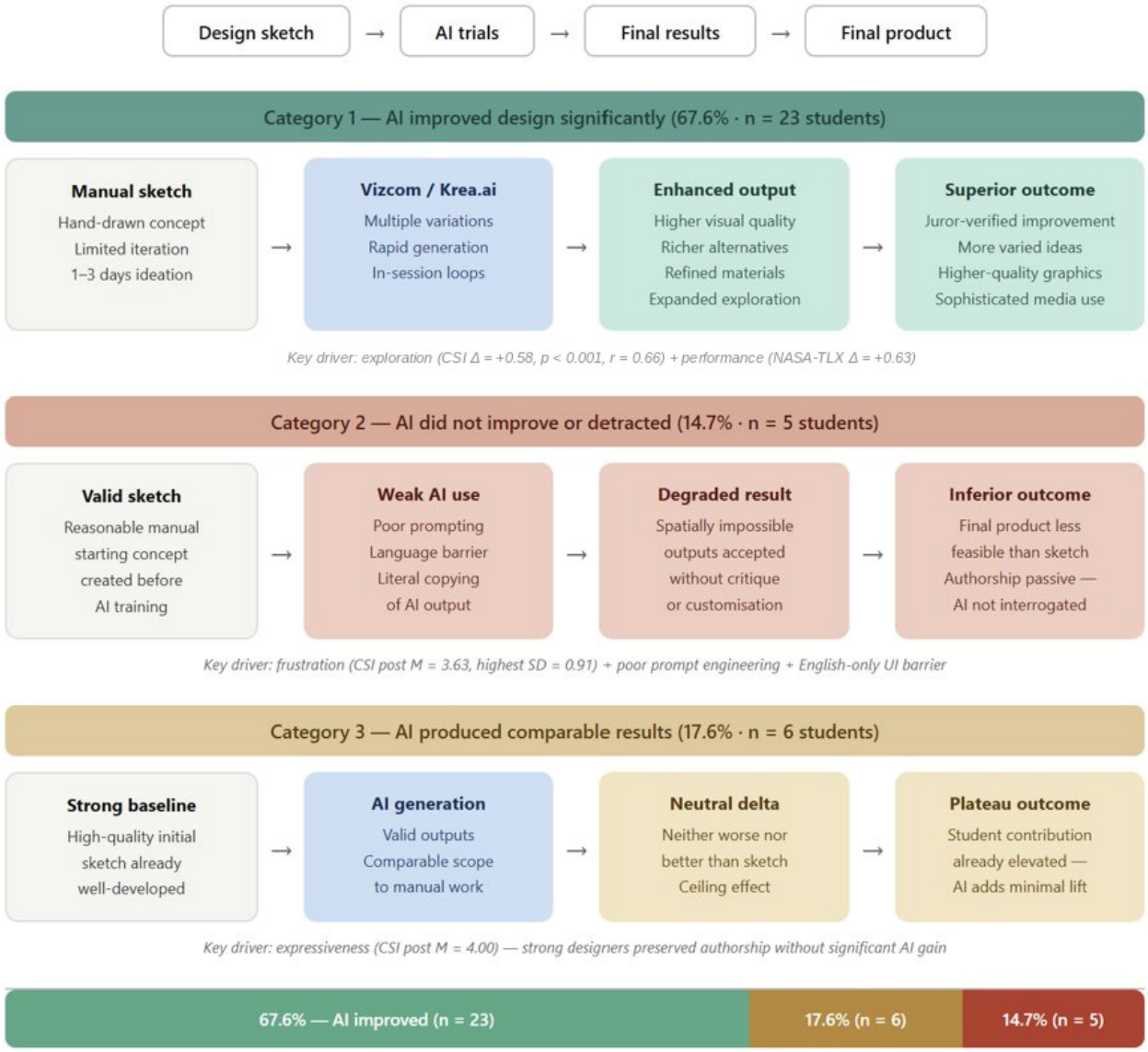


Figure 3. The workflow of outcomes framework.

Table 9 functions as the study's visual evidence layer, a portfolio display of actual student design work prepared across four progressive stages (Design Sketch → AI Trials → Final Results → Final Product) and categorized into the same three outcome groups used by the expert jurors in Figure 3 and Table 8. The analytical weight of Table 9 must be read alongside the surrounding quantitative findings, as the images illustrate the procedures behind each outcome category and give actual shape to the statistical patterns reported in Tables 6–8. The agreement between the two expert jurors was strong (Cohen's $\kappa = 0.76$), supporting the validity of the three-category classification system embodied in the table.

Category 1: AI Significantly Improved the Design (67.6%, n = 23). This is the major outcome and the one most tightly coupled with the study's strongest quantitative result: the large-effect-size gain in exploration ($Z = -3.42$, $p < 0.001$, $r = 0.66$). In these cases, the student's manual sketch served as a concept but was limited by time and hand-rendering ability. The AI trials phase, using Vizcom and Krea.ai, enabled the rapid generation of multiple design alternatives in a single session, compressing what was previously a 1–3 day ideation process. Jurors noted improvements in visual quality, variety of alternatives, graphical difficulty, and media value. The performance gain (NASA-TLX $\Delta = +0.63$) is

attributable almost entirely to this category, since juror-rated quality improvement was the primary driver of students' self-reported satisfaction with performance.

Category 2: Designer Sketch Better Than AI Output (14.7%, n = **5**). This category identifies the conditions under which AI integration fails. Jurors observed three failure mechanisms: students who did not prompt AI effectively produced generic or misaligned outputs, consistent with the frustration score (M = 3.63, SD = 0.91, the highest variance of all subscales); some students accepted AI outputs accurately without customization, falling into passive authorship; and the English-language interfaces of Vizcom and Krea.ai created a compounded cognitive problem for Arabic-speaking students who had to concurrently manage prompt engineering and cross-linguistic translation of design intent, a uniquely MENA specific equity barrier. These cases demonstrate that AI's value depends on prompt-writing skill, critical evaluation capacity, and access to linguistically appropriate tools.

Category 3: Similar Results Between Manual and AI-Assisted Work (17.6%, n = **6**). This outcome reveals a maximum effect: students who began with strong manual sketches experienced the least measurable lift from AI assistance. Jurors concluded that these students' initial work was already of sufficient quality that AI tools added minimal visual or conceptual improvement. This finding aligns with the high post-test expressiveness score (M = 4.00), indicating these students maintained creative ownership precisely because they were already operating with high design confidence. Consistent with Paananen et al. [15], authorship was most strongly preserved when students engaged critically with AI output rather than relying on it as a primary driver.

Triangulation with the Broader Study Results. The three categories in Table 9 serve as the humanizing, visible equivalent to the statistics in Tables 6–8. The significant exploration gain (CSI $\Delta = +0.58$, $p < 0.001$, r = 0.66) maps directly onto Category 1, where AI expanded idea generation beyond what manual sketching could achieve in a single session. The reductions in effort (NASA-TLX $\Delta = -0.56$) and temporal demand ($\Delta = -0.50$) are best understood through these same Category 1 samples, in which AI compressed multi-day ideation cycles into in-session loops. The high frustration variance (SD = 0.91) falls within the range of Category 2, indicating that some students successfully navigated prompt engineering, while others did not. Taken together, Table 9 confirms that generative AI is most beneficial as a conceptual phase during early ideation stages and is most likely to detract from quality when students engage with it passively, without the critical editorial judgment identified as the core pedagogical competency for responsible AI integration in design education.

5. Discussion

This study provides empirical support for the affordances of a generative AI tool in reimaging the undergraduate furniture design curriculum, along with notable limitations and learning considerations. Each subsequent section will briefly discuss its respective finding in the context of the related literature before discussing broader takeaways for design education and implications for Jordanian and MENA institutions.

5.1. Exploration as the Primary Creative Benefit

Our strongest result, a significant positive effect on exploration ($Z = -3.42$, $p < 0.001$, r = 0.66), also parallels findings from CSI-based research of AI-assistance in creative tasks. Remy et al. [12], Al Sheikh [39], and Al Sheikh, and Jazzar [40] benchmarked creative support tools integrated during HCI design, They found exploration to be the most commonly increased dimension across studies when AI tools were employed, arguing that breadth of ideation is a domain-general effect of generative AI tools rather than an artifact of any specific tool, platform, or design process. Paananen et al. [15] likewise reported

increased exploration with AI-assisted architectural ideation and found that expressiveness and authorship remained high, a finding reproduced here.

The majority report transitioning from a 1–3 day manual ideation process to an in-session concept-generation loop. A major instructional advantage of this acceleration is that students can iterate through rapid prototyping loops more closely aligned with how designers work in practice [2]. Theoretically, generative AI tools reduce the cost of generating many diverse design options and allow students to evaluate more alternatives than is feasible by hand. Importantly, the AI-assisted phase did not replace the original conceptual design process but instead extended and transformed existing manually developed concepts through accelerated visualization, exploratory iteration, and representational refinement. Consequently, the pedagogical value of generative AI in this study lies less in replacing students' conceptual agency and more in accelerating exploratory visualization cycles that enable students to critically compare, reject, and refine alternatives during early-stage furniture ideation.

Thematic analysis of open-ended responses ($\kappa = 0.81$) reveals that the exploration gain is not simply a function of increased output volume. Still, it reflects a qualitative shift in the students' relationship to early-stage design uncertainty. Students who made the greatest gains in exploration described using AI tools to "test directions before committing," effectively treating Vizcom and Krea.ai as low-cost externalization tools for speculative ideation. This aligns with the concept of cognitive offloading: AI absorbs the burden of visual representation, freeing students' working memory for evaluative and connective thinking. Importantly, several students noted that seeing AI-generated alternatives "showed me what I didn't want," indicating that AI-assisted exploration also functioned as a divergent-thinking prompt that sharpened students' own design preferences. This mechanism, which uses AI outputs as creative contrast rather than as creative substitutes, represents a pedagogically significant use pattern that warrants deliberate scaffolding in future AI-integrated studio curricula.

5.2. Cognitive Workload Reduction and Design Performance

Results from the NASA-TLX Likert-scale averages (Δ indicates a comparison) reveal significant reductions in perceived workload on the effort ($\Delta = -0.56$), temporal demand ($\Delta = -0.50$), and mental demand ($\Delta = -0.38$) subscales when using AI throughout the design process. This validates Chandrasekera et al.'s [5] findings that implementing generative AI drastically reduces mental effort during the early phases of the design process without sacrificing the quality of creative output. Improvements were also observed in perceived performance ($\Delta = +0.63$); learners self-reported being much more pleased with the design quality generated by AI, a finding supported by our juror analysis, which found that 67.6% of AI-generated artifacts were significantly better than their manual counterparts.

Taken together, these observations suggest that generative AI can facilitate cognitive scaffolding, enabling students to externalize the cognitive load associated with design work. Offloading shifts constraints away from bottlenecks at the ideation stage and onto students' ability to make judgment calls during editing and critique, where generative AI can't intervene [24]. There are educational consequences to this shift: if students are offloading the generation of ideas to AI, studio learning can gradually shift to focus more on critical thinking, contextual evaluation, and humane decision-making.

Workload, Performance, and Pedagogical Shift. Qualitative data extend and nuance the interpretation of cognitive scaffolding. Students did not simply describe a reduced workload in general terms; they identified the specific phases where relief was felt most acutely, particularly the transition from an empty page to a first visual form and the identification of viable material combinations. Informal interview data further revealed

that the perceived performance gain (NASA-TLX Δ = +0.63) was partly attributable to a comparison effect: students who viewed AI-assisted renders alongside their manual sketches reported feeling "surprised by how professional it looks" relative to their hand-drawn work, which may inflate the subjective performance gain beyond objective quality differences. Expert juror findings partially moderate this: while 67.6% of AI-assisted outputs were objectively rated as improved, 14.7% were rated as worse, indicating that perceived performance satisfaction does not uniformly correspond to actual design quality. This discrepancy is pedagogically significant and suggests that studio instruction should explicitly develop students' capacity to critically evaluate the visual quality of AI-generated work rather than automatically accepting it as superior.

5.3. Maintained Authorship and Expressiveness

The small but significant effect size for expressiveness (Δ = +0.21, $p = 0.205$) and the comparatively high post-test ownership score (M = 4.00) indicate that students still felt a sense of ownership over their AI-assisted creations. Kahraman et al. [1] reached a similar conclusion, finding that students' interpretation of AI-generated material was the primary determinant of design quality in AI studio projects, rather than prompt complexity. This also supports our initial idea that when students perceive AI as generative rather than independent, they maintain the skeptical faculties needed for true ownership. Uncertainty must be taught; if students are not taught to be critical of AI, they may approach AI-generated content passively, copying and pasting, which can lead to a decrease in design ability.

Authorship and the Active vs. Passive AI Engagement Distinction: Open-ended responses and interview data reveal a consistent pattern differentiating students who maintained strong authorship from those who did not; the former group consistently described a deliberate, two-stage engagement with AI, first generating a range of options, and then applying their own design criteria to select and transform those options. Representative statements included: "a tool I direct, not one that decides for me" and "I use it like I would a mood board, to trigger ideas, not to copy them." Students in the passive engagement group, by contrast, described adopting AI outputs more directly: "I used what it gave me because it looked better than what I could draw." This distinction between active and passive AI engagement is not captured by the post-test Likert scores alone and represents a critical qualitative finding for pedagogy: a high group-level expressiveness score (M = 4.00) may mask important variation in authorship quality. Future research should develop rubrics that distinguish between students who transform AI outputs and those who transplant them.

5.4. Persistent Frustration and Tool Limitations

While encouraging, these results must be tempered by the moderate frustration score on the post-test questionnaire (M = 3.63, SD = 0.91) and by the fact that 14.7% of AI-generated outputs did not improve students' initial concepts. Three primary sources of frustration were identified through qualitative analysis: the cognitive load of writing prompts; AI interfaces defaulting to English-language preconditions/platforms, which non-Anglophone students must overcome as a language barrier; and AI's limitations in translating complex constructional and spatial information. Both of the former points relate closely to Oppenlaender's [10] account of difficulties associated with prompt engineering and lend support to the argument that users face metalinguistic hurdles when writing effective prompts for creative purposes (especially the latter point concerning AI's limitations with more technical/construction-based information). Although Vizcom and Krea.ai contributed most strongly to rapid visual exploration and ideation expansion, ChatGPT primarily supported prompt refinement, conceptual clarification, and the development of material descriptions, suggesting that the strongest gains in exploration emerged from

integrating text- and image-based AI tools rather than from a single platform alone. This finding suggests that exploration gains emerged primarily through the complementary integration of conversational AI and image-generation systems rather than through the isolated use of a single tool.

Language ability presents a key equity barrier in the Jordanian context, particularly for students working to conceptualize furniture within the constraints of a furniture design program grounded in Arabic-language pedagogy, who are required to write creative prompts in English. This represents a unique metalinguistic requirement layered on the already significant cognitive load of writing an effective prompt. Platforms in the future with more robust multilingual UI would significantly reduce this equity barrier. Within the broader MENA region, this issue extends beyond Jordan and reflects a structural imbalance in the development of AI platforms that primarily privilege English-language interaction. Many Arabic-speaking students must simultaneously manage design ideation, translation of technical terminology, and prompt optimization, creating a dual cognitive burden that may disadvantage non-native English speakers. Furthermore, AI-generated outputs frequently reflect Western-centric aesthetic references that may not adequately represent regional cultural identities, vernacular furniture traditions, or local design vocabularies. These findings suggest that future AI platforms used in design education should incorporate Arabic-language prompting systems, culturally adaptive datasets, and localized interface support to ensure equitable participation across MENA educational environments.

Frustration Typology and Pedagogical Implications. Thematic analysis of post-test open-ended responses allows the three frustration sources identified in Section 5.4 to be further differentiated into acute and structural barriers. Prompt engineering difficulty was experienced as an acute barrier, one that students reported partially overcoming over the four-week study as they became more familiar with effective prompt phrasing. Representative statements included "the second session was easier because I knew what words to use." The English-language interface barrier, by contrast, was experienced as a structural and unresolvable constraint during the study period: students described it as a persistent cognitive tax that required them to mentally translate design concepts into English before articulating them as prompts and then re-evaluating the outputs through an Arabic-language design lens. This double-translation burden was identified in informal interviews as the single most disruptive factor for students whose first language was Arabic. The third frustration, AI's inability to render intricate constructional details such as joinery, upholstery attachment, and precise dimensional relationships, was experienced as an inherent platform limitation rather than a learnable skill gap. These findings suggest that frustration in AI-assisted design education is not monolithic: pedagogical interventions can address prompt engineering and language barriers through training and tool selection, but students and educators must also develop realistic expectations about AI's technical limitations in detail-specific design domains.

5.5. Collaboration and the Limits of Prompt-Based Interaction

The lack of a significant difference in collaboration (Δ = +0.06, p = 0.713) can be explained by acknowledging that today's generative AI applications are built around solo interactions between humans and AI. AI-mediated peer learning exchanges, frequently observed in productive studio-centered design learning, cannot easily occur without being designed for. Thus, when embedding generative AI into studio learning experiences moving forward, opportunities should be consciously built in for collaboration around AI-mediated activities (e.g., group-generated prompts, peer review of AI-generated options, group iteration on AI-supported designs).

5.6. Contextual Implications for Jordanian and MENA Design Education

Conducted in Jordanian contexts, our study enriches the Western- and East Asia-dominated literature on generative AI by contributing empirical findings from the MENA region. The positive predisposition Jordanian design students demonstrated towards employing AI tools (evidenced by positive CSI scores in the pre-test and improvement in all CSI dimensions in the post-test) suggests that cultural and/or language differences are not insurmountable barriers to AI integration in design education contexts with the necessary language and instructional accommodations [1,2]. The language differences noted in the qualitative data suggest that the unique circumstances of MENA classrooms should inform bespoke implementation strategies that have not yet been discussed in the current literature. Practical curricular implementation could include dedicated prompt-engineering workshops in first-year digital design courses, AI-supported ideation exercises in furniture and interior studios, and reflective assignments that require students to document how AI influenced their design decisions. Design programs may also integrate bilingual AI exercises where students develop prompts in both Arabic and English to strengthen technical vocabulary and reduce linguistic dependency barriers. Additionally, faculty-led critique sessions comparing manual and AI-assisted outputs could help students develop critical evaluation skills and maintain awareness of authorship. Recent research on AI-assisted interior workflows and educational integration further confirms these pedagogical transformations [13–16].

6. Conclusions

Findings from this work provide quantitatively substantiated evidence of the affordances of generative AI tools in revolutionizing undergraduate furniture design coursework, with a focus on the Jordanian educational setting. Performance differences between conditions yielded reliable, meaningful positive effects. Integrating Vizcom, Krea.ai 1, and ChatGPT resulted in statistically significant improvements on the CSI exploration factor ($Z = -3.42$, $p < 0.001$, $r = 0.66$) and meaningful improvements on the workflow factor ($Z = -1.97$, $p = 0.049$, $r = 0.38$). Participants also showed reliable positive directions of improvement for creativity, expressiveness, and enjoyment [9]. Juror ratings backed up these findings objectively. In total, 67.6% of AI-assisted student projects earned higher ratings than non-AI-assisted projects. These findings have led to the conclusion that, when carefully directed, generative AI functions optimally as an extension of the student creator. It can provide students with methods to broaden their ideation, reduce mental fatigue, and produce more high-quality designs, all while maintaining student authorship. The findings therefore support interpreting generative AI primarily as a collaborative augmentation and iterative co-design support system rather than as an autonomous replacement for human creativity. These reframing positions AI-assisted furniture design education within a human-in-the-loop pedagogical model where students remain responsible for contextual judgment, cultural interpretation, and critical evaluation of generated outputs.

Frustrations persist around prompt engineering, language access, and output specificity. Continued irrelevance in collaboration also suggests that sessions between a human and AI, with prompts, will not naturally create the social learning found in studio pedagogy.

7. Recommendations

1. Use AI the most during the ideation phase: Use AI only for early exploratory tasks. Use traditional methods for the final phase of development, such as space planning, construction, and material specification.

2. Take a class on prompt engineering: Studios should have specific classes on how to craft prompts. This would include teaching linguistic skills to ESL students as well as how to iterate on prompts. I feel every studio utilizing AI should take these courses.
3. Plan opportunities for sharing how to use AI tools: Studio activities should incorporate students using AI in conjunction with one another. This could look like coming up with prompts as a group or sharing what AI can do with one another.
4. Institute policy around authorship/attribution: Departments might institute policies where students must keep track of how they use AI tools in a design and be clear about what elements were made by AI vs. the students. This creates a sense of responsibility and normalizes treating AI as any other design tool [2].
5. Lobby for multilingual and culturally inclusive AI platforms: Institutions teaching design in non-Anglophone environments should work alongside developers of AI tools to advocate for comprehensive multilingual user interface support and culturally inclusive training datasets.

This study suggests several areas for future work. Longitudinal measurement of AI skill acquisition over several semesters would verify whether the gains from exploration and efficiency endure over time with prolonged tool use. Administering the NASA-TLX immediately before and after the pre- and post-tests would enable stronger comparisons between the two workload measures, rather than relying solely on face validity, given the asymmetry in measurement [4]. External validity may be enhanced through investigating other universities in Jordan and the Arab world, as well as international universities. Further work is necessary as generative artificial intelligence design tools develop. Interior and furniture design curricula should prepare students to use AI as a supplement to their design practices by developing critical, ethical, and locally conscious AI frameworks that enhance human creativity.

Funding: This research received no external funding.

Institutional Review Board Statement: Ethical review and approval were waived for this study due to not involving humans or animals.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Skipworth, S.; Chatteur, F.; Maher, M.L.; Jan, T. The acceptance of artificial intelligence (AI) by an Australian interior design and design education community. *Int. J. Des. Creat. Innov.* **2025**, *13*, 93–113. [CrossRef]
2. Huh, M.B. Students' perceptions of generative AI image tools in architecture and interior architecture education. *Educ. Sci.* **2025**, *15*, 1160. [CrossRef]
3. Demirbaş, Ö.O.; Demirkan, H. Learning styles of design students and the relationship of academic performance and gender in design education. *Learn. Instr.* **2007**, *17*, 345–359. [CrossRef]
4. Hart, S.G.; Staveland, L.E. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Human Mental Workload*; Hancock, P.A., Meshkati, N., Eds.; North-Holland: Amsterdam, The Netherlands, 1988; pp. 139–183. [CrossRef]
5. Chandrasekera, T.; Hosseini, Z.; Perera, U. Can artificial intelligence support creativity in early design processes? *Int. J. Archit. Comput.* **2024**, *23*, 112–136. [CrossRef]
6. Li, Z.; Sayuti, N.A.A.; Jinfeng, W. The impact of AIGC technology on furniture design education. *Malays. J. Soc. Sci. Humanit.* **2025**, *10*, e003173. [CrossRef]
7. Shneiderman, B. *Human-Centered AI*; Oxford University Press: Oxford, UK, 2022.
8. Liu, Y.; Xu, B.; Feng, J.; Wu, P. Analysis of the application of generative artificial intelligence in interior design education. *Ain Shams Eng. J.* **2025**, *16*, 103757. [CrossRef]

9. Wang, K.; Cheng, L.; Yin, M.; Zhang, K.; Wang, R.; Zhang, M.; Sun, R. Evolutionary Game Theory in Energy Storage Systems: A Systematic Review of Collaborative Decision-Making, Operational Strategies, and Coordination Mechanisms for Renewable Energy Integration. *Sustainability* **2025**, *17*, 7400. [CrossRef]
10. Cherry, E.; Latulipe, C. Quantifying the creativity support of digital tools through the creativity support index. *ACM Trans. Comput. Hum. Interact.* **2014**, *21*, 1–25. [CrossRef]
11. Oppenlaender, J. The creativity of text-to-image generation. In *Proceedings of the 25th International Academic Mindtrek Conference, Tampere, Finland, 16–18 November 2022*; ACM: New York, NY, USA, 2022; pp. 192–202. [CrossRef]
12. Remy, C.; MacDonald Vermeulen, L.; Frich, J.; Biskjaer, M.M.; Dalsgaard, P. Evaluating creativity support tools in HCI research. In Proceedings of the 2020 ACM Designing Interactive Systems Conference, Eindhoven, The Netherlands, 6–10 July 2020; pp. 457–476. [CrossRef]
13. Braun, V.; Clarke, V. Using thematic analysis in psychology. *Qual. Res. Psychol.* **2006**, *3*, 77–101. [CrossRef]
14. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed.; Lawrence Erlbaum Associates: Hillsdale, NJ, USA, 1988.
15. Paananen, V.; Oppenlaender, J.; Visuri, A. Using text-to-image generation for architectural ideation. *Int. J. Archit. Comput.* **2024**, *22*, 458–474. [CrossRef]
16. Hart, S.G. NASA-Task Load Index (NASA-TLX); 20 years later. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2006**, *50*, 904–908. [CrossRef]
17. Kahraman, M.; Şekerci, Y.; Develier, M.; Koyuncu, F. Integrating Artificial Intelligence in Interior Design Education: Concept Development. *J. Comput. Des.* **2024**, *5*, 31–60. [CrossRef]
18. Li, C.; Zhang, T.; Du, X.; Zhang, Y.; Xie, H. Generative AI models for different steps in architectural design: A literature review. *Front. Archit. Res.* **2025**, *14*, 759–783. [CrossRef]
19. Liu, Y. Research on the Application of Artificial Intelligence in Interior Design. *Int. J. Sci. Eng. Appl.* **2024**, *13*, 24–29. [CrossRef]
20. He, Z.; Li, X.; Fan, L.; Wang, H. Revamping Interior Design Workflow Through Generative Artificial Intelligence. In *International Conference on Human-Computer Interaction*; Springer: Cham, Switzerland, 2023; pp. 607–613. [CrossRef]
21. Chen, J.; Shao, Z.; Hu, B. Generating Interior Design from Text: A New Diffusion Model-Based Method for Efficient Creative Design. *Buildings* **2023**, *13*, 1861. [CrossRef]
22. Hong, M.K.; Hakimi, S.; Chen, Y.Y.; Toyoda, H.; Wu, C.; Klenk, M. Generative AI for product design: Getting the right design and the design right. *arXiv* **2023**, arXiv:2306.01217. [CrossRef]
23. Dezeen. Philippe Starck, Kartell and Autodesk Unveil "World's First Production Chair Designed with Artificial Intelligence". 11 April 2019. Available online: https://www.dezeen.com/2019/04/11/ai-chair-philippe-starck-kartell-autodesk-artificial-intelligence-video/ (accessed on 1 April 2026).
24. Rane, N.; Choudhary, S.; Rane, J. Enhancing sustainable construction materials through the integration of generative artificial intelligence, such as ChatGPT. *Sustain. Clean Build.* **2026**, *1*, 98–122. [CrossRef]
25. Rane, N. Potential Role and Challenges of ChatGPT and Similar Generative Artificial Intelligence in Architectural Engineering. *SSRN Electron. J.* **2024**, *4*, 22–47. [CrossRef]
26. Kaivo, M.; Cuartielles, D.; Spikol, D. Designing activities and tools to support university students' creative and collaborative exploration of physical computing. *Networked Learn. Conf.* **2024**, *12*, 55–64. [CrossRef]
27. Sano, S.; Yamada, S. AI-Assisted Design Concept Exploration Through Character Space Construction. *Front. Psychol.* **2022**, *12*, 819237. [CrossRef] [PubMed]
28. Wang, Z.; Tan, Z.; Ma, Y. Combinediff: A GenAI creative support tool for image combination exploration. *J. Eng. Des.* **2024**, *36*, 731–757. [CrossRef]
29. Medel-Vera, C.; Britton, S.; Gates, W.F. An exploration of the role of generative AI in fostering creativity in architectural learning environments. *Comput. Educ. Artif. Intell.* **2025**, *9*, 100501. [CrossRef]
30. Shao, Z.; Chen, J.; Zeng, H.; Hu, W.; Xu, Q.; Zhang, Y. A new approach to interior design: Generating creative interior design videos of various design styles from indoor texture-free 3D models. *Buildings* **2024**, *14*, 1528. [CrossRef]
31. Braarud, P.O. An efficient screening technique for acceptable mental workload based on the NASA Task Load Index—Development and application to control room validation. *Int. J. Ind. Ergon.* **2020**, *76*, 102904. [CrossRef]
32. Barajas-Bustillos, M.A.; Maldonado-Macías, A.; Serrano-Rosa, M.; Hernández-Arellano, J.; Llamas-Alonso, L.; Balderrama-Armendariz, O. Impact of experience on the sensitivity, acceptability, and intrusive of two subjective mental workload techniques: The NASA TLX and workload profile. *Work* **2023**, *75*, 1265–1275. [CrossRef] [PubMed]
33. Hussein, G.K. Improving design efficiency using artificial intelligence: A study on the role of artificial intelligence in streamlining the interior design process. *Int. Des. J.* **2023**, *13*, 255–270. [CrossRef]
34. Dewit, I.; Latulipe, C.; Dams, F.; Jacoby, A. Using the creativity support index to evaluate a product-service system design toolkit. *J. Des. Res.* **2021**, *18*, 434–457. [CrossRef]
35. Creswell, J.W.; Creswell, J.D. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 5th ed.; SAGE Publications: Thousand Oaks, CA, USA, 2018.

36. IBM Corp. *IBM SPSS Statistics for Windows, Version 28.0*; IBM Corp.: Armonk, NY, USA, 2021.
37. Anantrasirichai, N.; Bull, D. Artificial intelligence in the creative industries: A review. *Artif. Intell. Rev.* **2022**, *55*, 589–656. [CrossRef]
38. Lin, Y.C.; Chien, C.Y.; Hou, K.H.; Chen, H.Y.; Kuo, P.C. AIDED: Augmenting interior design with human experience data for designer-AI co-design. In Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems, Barcelona, Spain, 13–17 April 2026; pp. 1–29. [CrossRef]
39. Al Sheikh, H.M. Design thinking and quality management in architecture and design faculties: From accountability to empowerment and creativity. *Creat. Stud.* **2026**, *19*, 51–65. [CrossRef]
40. Al Sheikh, H.M.; Hamza Al Jazzar, F. Mind maps in design education: Enhancing cognitive and strategic thinking–A case study review from Al-Ahliyya Amman University. *Des. J.* **2026**, *29*, 428–452. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.