Comparative Study of Generative AI and Traditional Tools for Evaluating Creativity and Efficiency in UI/UX Design

Iqra Shahzad, Marriam Daud, Wajiha Mughal
Department of Software Engineering, National University of Modern Languages, Islamabad 44000, Pakistan
*Correspondence: marriam.daud@numl.edu.pk

Citation | Shahzad. I, Daud. M, Mughal. W, "Comparative Study of Generative AI and Traditional Tools for Evaluating Creativity and Efficiency in UI/UX Design", IJIST, Vol. 8 Issue. 2 pp 924-940, May 2026
Received | March 28, 2026 **Revised |** April 30, 2026 **Accepted |** May 05, 2026 **Published |** May 11, 2026.

In the modern software industry, the creation of user interface (UI) and user experience (UX) design is an essential activity that is increasingly adopting generative artificial intelligence (AI) tools. With this growing trend, it is crucial to examine the novelty and creativity of AI-generated design outputs in comparison to those produced by traditional workflows without any AI support. The goal of this study is to quantitatively compare the cognitive load, creativity, and task completion time of UI/UX design outputs created using the traditional design tool named Figma and the generative AI tool named Galileo AI. The NASA task load index (NASA-TLX), creative product semantic scale (CPSS), and number of minutes taken to complete a design task are used to quantify cognitive load, creativity, and task completion time, respectively. To compare these three metrics between AI and traditional groups, an experiment is conducted with novice and expert designers, separately. The results reveal that the generative AI tool significantly reduces both the cognitive load of designers and the time required to create UI/UX designs. For instance, independent samples t-tests show that the generative AI tool results in lower cognitive load scores than the traditional tool for both novices (2.22 ± 0.53 vs. 3.25 ± 0.70) and experts (2.42 ± 0.64 vs. 3.79 ± 0.52). Similarly, independent samples t-tests results indicate that the AI tool reduces average task completion time for both novices (24.38 vs. 42.69) and experts (23.78 vs. 48.44) compared to the traditional tool. On the other hand, the traditional tool consistently outperforms the generative AI tool in terms of creativity and innovation. For instance, independent samples t-tests results show that the traditional design tool yields higher creativity scores than the generative AI tool for both novices (3.94 ± 0.66 vs. 3.27 ± 0.74) and experts (4.17 ± 0.39 vs. 3.37 ± 0.44). Overall, the AI tool excels at enhancing efficiency; however, it still faces challenges in replicating the depth of human creativity and emotional understanding that are essential for creating engaging UI/UX designs.

Keywords: Cognitive Load; Creativity; Generative AI tool; UI/UX Design Creation; Workflow Efficiency.

Introduction:

In today's software industry, user interface (UI) and user experience (UX) design play an integral role in crafting effective and engaging software systems. The true essence of UI/UX design lies in its emphasis on usability, efficiency, retainability, usefulness, novelty, creativity, and meaningfulness [1]. When these qualities are missing, the consequences can be severe, ranging from user dissatisfaction to significant financial losses, and in some cases, even complete project failure [1].

Traditionally, designers rely on tools such as Figma [2], Adobe XD [3], or Sketch [4] to create interactive prototypes, a process that heavily draws on their skills, creativity, and decision-making abilities. The rapid adoption of generative artificial intelligence (AI) tools, such as Midjourney [5], DALL·E 2 [6], Stable Diffusion [7], and Galileo AI [8], has introduced new possibilities for automating design tasks. While the use of these tools in UI/UX design workflows is rapidly expanding, there is a critical need to quantitatively assess their impact on two fundamental pillars of UI/UX design - creativity and efficiency. Creativity in UI/UX design is understood as the ability to generate solutions that are not only novel but also useful and appropriate [9][10]. Efficiency, on the other hand, refers to delivering high-quality outcomes while minimizing task completion time, cognitive load [11], and the use of resources.

Recent studies [12][13] found that generative AI significantly supports creativity by enhancing idea generation, content creation, and design automation in UX workflows. However, they also highlight challenges such as ethical concerns, bias in generated outputs, lack of transparency, questions of authorship, overreliance on AI systems, and insufficient integration of human-centred design principles. [14] Conducted an exploratory study with 31 students and found that the AI4Design system supports students in improving conceptual clarity and visual outcomes, indicating increased creativity. Another study [15] found that generative AI tools improved collaboration, idea generation, and productivity in UX design courses. [16] found that generative AI is not only a tool for automation but also functions as a co-creator that actively shapes the design process.

In recent years, many studies [17][18][19][20][21][22][23] have highlighted that while generative AI tools enhance the efficiency of UI/UX design creation, they also pose risks and challenges to the design industry, particularly in terms of reducing creativity. However, these studies were limited as they mainly relied on assessments based on designers' perspectives and small sample sizes, instead of objectively assessing these qualities through UI/UX design outputs.

In this study, we quantitatively compared the cognitive load, creativity, and task completion time between UI/UX design outputs created using a generative AI tool named Galileo AI and the traditional tool named Figma. For this purpose, a between-participants experiment is performed with 26 senior students (designated as novices) as well as 18 UI/UX design professionals (designated as experts), separately. In both categories (novices and experts), the participants are randomly divided into two equal groups. One group (designated as traditional) used Figma to create a particular design output, while the other group (designated as AI) used Galileo AI.

Research Objectives:

This study has the following main objectives:

To evaluate and compare the cognitive load experienced by designers when performing UI/UX design tasks with a generative AI tool versus a traditional design tool.

To compare the creativity of design outputs produced using a generative AI tool and a traditional design tool.

To measure the difference in task completion time between a generative AI tool and a traditional UI/UX design tool.

To utilize the NASA-TLX, CPSS, and the number of minutes taken to complete a design task to measure cognitive load, creativity, and task completion time, respectively.

To conduct a between-participants experiment to collect the cognitive load, creativity, and task completion time scores of a generative AI design tool versus a traditional design tool for different experience levels, including novice designers and expert UI/UX professionals, separately.

To prepare two datasets, one for novices and one for experts.

To perform an independent samples t-test for both novices and experts to compare cognitive load, creativity, and task completion time results between a generative AI tool group and a traditional tool group.

Research Questions:

Through this study, we attempt to answer the following research questions (RQs).

RQ1: How does the generative AI tool affect the cognitive load experienced by UI/UX designers compared to the traditional tool?

RQ2: How does the generative AI tool impact creativity and innovation in UI/UX designs compared to the traditional tool?

RQ3: How does the generative AI tool affect the time required to complete UI/UX design tasks compared to the traditional tool?

The rest of this paper is structured as follows. Section 2 presents a comparative summary of the related studies on creativity and efficiency in UI/UX Design. Section 3 outlines our research methodology. Section 4 provides the results of the comparative analysis between AI and traditional tool groups. Section 5 offers a discussion of threats to the validity of this study and corresponding mitigation strategies. Section 6 states the implications of this study. Section 7 presents our recommendations for practitioners and researchers. Finally, Section 8 discusses our main contributions and highlights directions for future research.

Related Work:

Table 1 presents a comparative summary between this study and the related studies referenced in this section by highlighting parameters (domain, number of participants, research type, assessment methods/tools, and limitations) of these studies.

[17] Conducted a survey based on categories of the SPACE framework with 72 professionals to investigate the impact of generative AI technologies on creativity, efficiency, usability, and ethical concerns related to creating UI/UX designs of mobile applications. The results of this survey showed that generative AI tools improve the efficiency of creating UI/UX design solutions, yet they tend to decrease the creativity, diversity, and originality of the UI/UX design solutions. However, this study is limited by a relatively small sample size. Moreover, it was relying on UI/UX designers' self-assessments rather than objective evaluations based on actual design outputs.

[24] Evaluated differences in cognitive load and creativity between two groups of 20 design students. Both groups were tasked to design urban furniture through sketches and 3D renderings. One group used a generative AI tool (ChatGPT) for conceptualization for conceptualization, while the other worked without AI assistance. The cognitive load and creativity were assessed using the NASA Task Load Index (NASA-TLX) [25] and creative product semantic scale (CPSS) [26], respectively. The results of this study showed that the AI-assisted group achieved enhanced creativity and lower cognitive load scores compared to the baseline group. However, the absence of professional designers in this study may limit the applicability of these results to real-world settings.

[18] Conducted two surveys to investigate the perspectives of UI/UX design professionals regarding the role of generative AI tools in the UI/UX design workflows. While most professionals supported AI tools for improving their effectiveness and performance in design tasks, they also expressed concerns about the lack of innovation and creativity in

Table 1. Comparison summary of the related work

Sr. #	Author(s) (Year)	Domain	Participants (Research Type)	Assessment Methods and Tools	Limitations
1	Shahab (2024)	UI/UX design for mobile apps	72 Professionals (Survey)	Creativity and efficiency using the SPACE framework	Small sample size Based on designers' self-assessments No metrics-based assessment
2	Chandrasekera *et al.* (2024)	Interior design and architecture	40 Students (Quasi-experimental design)	Creativity using CPSS and cognitive load using NASA TLX	Not involving experienced design professionals
3	Chaudhry (2024)	UI/UX design	37 Professionals (Survey)	Designers' attitude towards AI tools using close-ended Likert scale questions Descriptive and thematic analysis	Small sample size No metrics-based assessment
4	Li *et al.* (2024)	UI/UX design	20 Professionals (Semi-structured interview)	Creativity using descriptive and thematic analysis	Small sample size No metrics-based assessment
5	Wadinambiarachchi *et al.* (2024)	Human computer interaction	60 Students (Experiment, Questionnaire, and Interview)	Creativity using human design fixation, fluency, variety, and originality	Not involving experienced design professionals
6	Chen *et al.* (2025)	Human computer interaction	10 Novice designers (Experiment and Semi-structured interview)	Generative AI design fixation by novelty	Small sample size Not involving experienced design professionals
7	Obanya (2025)	UI/UX design	30 Professionals (Survey), 5 Professionals (Semi-structured interviews), and 3 Real-world case studies	Creativity, efficiency, and usability using close-ended Likert scale questions Descriptive and thematic analysis	Small sample size Based on designer's self-reported responses No objective assessment of efficiency, usability, and creativity
8	Wu *et al.* (2026)	Human computer interaction	22 Students and 12 Professionals (Experiment, Survey, and Interviews)	4P model of creativity, Creative Self-Efficacy and Consensual	Participants used different generative AI tools

				Assessment Technique	
9	Shahzad *et al.* (2026) Current Study	UI/UX design	44 Participants 26 Students and 18 Professionals (Experiment and Questionnaire)	Creativity using CPSS, cognitive load using NASA TLX and task completion time	Based on one generative AI tool

AI-generated design outputs. However, this study relies primarily on descriptive and thematic analysis and is based on a relatively small sample size.

[19] Conducted semi-structured interviews with 20 UX designers to explore the impact of generative AI on UX design practices. Their findings indicate that experienced designers maintain creativity and originality when using generative AI tools, due to their strong foundational design skills. However, junior designers' greater dependence on AI tools may hinder the development of original ideas and weaken their creative problem-solving abilities. The key limitations of this study include a relatively small sample size and the lack of metrics-based assessment of actual design outputs.

[20] Quantitatively assessed design creativity using four criteria, including design fixation, variety, fluency, and originality. They asked participants to create sketches based on a given design brief under three different workflows: with access to an image-generating AI tool (Midjourney), with access to Google Image Search, and without any AI support (baseline). The results indicated that exposure to AI-generated images led to increased design fixation scores and reduced levels of fluency, variety, and originality compared to other conditions.

[21] Conducted an experiment with 10 novice designers to assess the novelty and creativity of design outcomes produced by generative AI tools. They revealed that design outputs produced through generative AI tools often struggle to achieve novelty or creativity due to generative AI tool's design fixation. The repetition or similarities in the design outputs generated by these tools occur because such tools depend extensively on existing data and learned patterns, which limits their capacity to produce diverse and innovative creative designs. They further mentioned that even when AI-generated designs are refined by humans, the final outputs may still lack novelty and diversity due to the influence of human design fixation.

[22] Examined the impact of AI-powered tools - Adobe Firefly, Figma AI, and Wizard - on design creativity, usability, and efficiency through 5 semi-structured interviews, an online survey of 30 UI/UX designers, and analysis of three real-world case studies. This study indicated that these tools enhance design efficiency and usability. However, many professionals expressed concerns about reduced creativity and increasing design homogenization. A key limitation of this study is that the findings relied solely on UI/UX designers' self-reported responses, without incorporating objective evaluations of efficiency, usability, and creativity.

[23] Conducted a controlled experiment to investigate how generative AI tools affect design creativity across the four dimensions of the 4P model (person, process, product, and press). The study involved 33 participants including 21 students and 12 professionals who created household appliance designs using both conventional methods and generative AI tools in two separate sessions. The findings indicate that generative AI enhances design efficiency while reducing the originality of the outcomes.

Novelty:

A critical analysis of the related studies reveals several important research gaps in the existing literature on generative AI in design. Most prior studies have focused on only a single evaluation factor, such as creativity, usability, productivity, or cognitive workload, rather than examining multiple dimensions together. Many of these studies investigated design domains other than UI/UX design, including graphic design, industrial design, architectural design, or general creative tasks. As a result, their findings cannot be directly generalized to the UI/UX design context.

Furthermore, existing related studies rarely provide a comprehensive evaluation incorporating cognitive load, creativity, and task completion time simultaneously within the same experimental framework. Another major limitation is that most previous studies considered only a single participant category, either novice designers or professional designers, without comparing the impact of generative AI tools across different expertise levels. Consequently, little is known about whether generative AI benefits beginners and experts differently in UI/UX design tasks.

Therefore, the current study attempts to fill these gaps by providing an empirical and quantitative comparison between the generative AI design tool and the traditional UI/UX design tool from both efficiency and creativity perspectives. Unlike prior studies, this research simultaneously evaluates cognitive load, creativity, and task completion time while involving both novice and professional designers.

Research Methodology:

We conducted a between-participants experiment to evaluate the cognitive load, creativity, and task completion time of design solutions generated using generative AI compared to human-generated solutions using a traditional UI/UX design tool. The independent variable is the type of UI/UX design tool: the generative AI-based UI/UX design tool (Galileo AI) and the traditional UI/UX design tool (Figma). The dependent variables are cognitive load, creativity, and task completion time. The experiment is performed in a controlled laboratory setting. We carefully adhered to ethical practices throughout the study when working with human participants. All participants are clearly informed about the study's purpose and procedures prior to the commencement of this study. In this regard, participants are educated about their entitlement and discontinue engagement with the study at any point without experiencing any consequences. Ethical treatment of participants is preserved by acquiring informed consent before the study begins.

Tools:

In this study, Figma and Galileo AI are used to create UI/UX designs. Figma is selected because it is one of the most widely used traditional design tools in the UI/UX industry to create interactive prototypes for mobile applications, websites, and other digital products. It represents the conventional approach to UI/UX design, where outcomes heavily depend on the designer's creativity.

On the other hand, Galileo AI is a generative AI tool that helps to create UI designs for mobile applications and websites using techniques such as text prompts or image-based inputs. Recently, Galileo AI was acquired by Google and has since been rebranded as Stitch [8].

Assessment Methods:

We employed the NASA Task Load Index (NASA-TLX) to measure the cognitive load of participants. NASA-TLX is a widely recognized instrument for assessing perceived workload across six dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration [25]. The rationale for selecting NASA-TLX is that it offers a comprehensive assessment of cognitive workload in the design context and has been widely applied in previous HCI and design-related studies.

We measured the creativity of UI/UX design outputs using the Creative Product Semantic Scale (CPSS). This scale evaluates creativity based on four important pillars – Novelty (originality and uniqueness of the design outcome), Resolution (the usefulness and practicality of the design outcome), Elaboration, and Synthesis (the degree of detail and the integration of elements within the design outcome along with aesthetic quality and overall elegance of the design outcome) [26]. Task completion time is measured as the number of minutes taken to complete a particular UI/UX design task.

Procedure:

Figure 1 illustrates a visual overview of our research methodology which consists of the following sequential steps:

Phase 1: Initiation

Participant Recruitment: This study sample comprised a total of 44 participants. These participants belonged to two categories: 26 senior students (final-year undergraduate students) from a public university in Islamabad, Pakistan, and 18 UI/UX design professionals, each with at least one year of experience in the UI/UX design field. These professionals were recruited through the UX community and LinkedIn and were employed at software organizations based in Islamabad, Pakistan. The participants are selected based on specific eligibility criteria, including their year of study, prior experience with design tasks, and familiarity with design software. All participants possessed foundational design knowledge and proficiency in Figma. Additionally, while they had prior experience utilizing various generative AI tools for UI/UX design, none had previously used Galileo AI. The sample size of 44 participants is determined based on feasibility considerations and is consistent with similar studies in the literature.

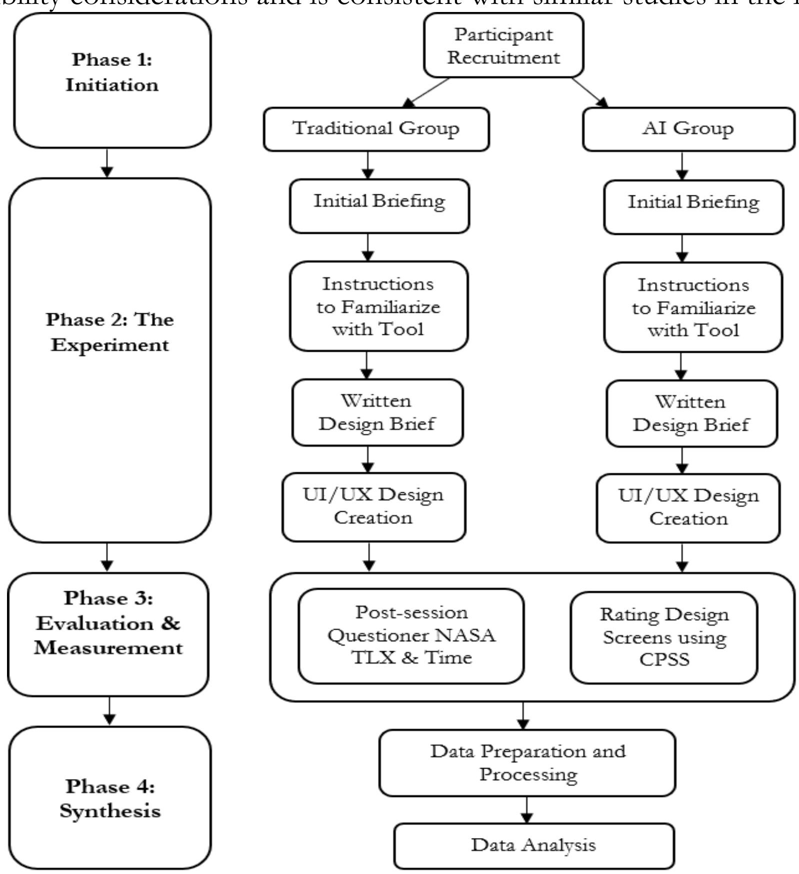

Figure 1. Our Research Methodology

The Split:

We randomly divided both types of participants into two groups i.e. AI and traditional. In the case of senior students (designated as novices), each group consisted of 13 senior students. On the other hand, for UI/UX design professionals (designated as experts), each group included 9 experts. In both categories (novices and experts), the traditional group developed the design solutions using the traditional UI/UX design tool i.e. Figma, while the AI group generated the design solutions using the generative AI-based UI/UX design tool i.e. Galileo AI.

Phase 2: The Experiment

Both groups went through an identical sequence of events, but using different tools:

Initial Briefing:

Participants are introduced to the study's goals.

Instructions to Familiarize with Tool:

Participants are given instructions to familiarize them with their assigned design tool (e.g., Figma for the traditional group vs. Galileo AI for the AI group).

Written Design Brief:

Both groups (AI and traditional) then received the identical design brief to create a "Patient Overview" screen for a healthcare application. The written design brief provided to participants stated: "Your task is to design a 'Patient Overview' screen for a healthcare application. The screen should include a top navigation bar, a patient information card, a health summary section, and an upcoming appointments section."

UI/UX Design Creation:

Participants in both groups proceeded to design the user interface screens based on the brief. Upon completion, all design outputs (screens) were collected and assigned a unique identifier. Precisely, among the novices, a total of 26 design screens are collected, 13 from the AI group and 13 from the traditional group. For the experts, 18 design screens are obtained, with 9 produced in each group.

Phase 3: Evaluation & Measurement

Once the designs are complete, data is collected using three metrics:

Post-session Questioner NASA TLX & Time:

Each group was asked to fill out a post-session questionnaire containing items from the NASA TLX to assess their cognitive load, as well as a question regarding task completion time (recorded using the timestamp).

Rating Design Screens using CPSS:

To objectively assess novelty and creativity, the UI/UX design screens are rated by two UI/UX design professionals using items from the CPSS.

Phase 4: Synthesis:

Data Preparation and Processing:

We used a 5-point Likert scale to compute the NASA TLX and CPSS scores. For instance, using NASA TLX, each participant's responses across the six items (questions) are collected. Subsequently, the responses for the six items are averaged to calculate the final NASA TLX score for each participant. Similarly, the final CPSS score for each participant is calculated by averaging six related Likert items (questions).

The individual Likert scale items are ordinal in nature (i.e. ranked categories). However, the NASA TLX and CPSS scores represent interval data since they were calculated by averaging multiple related Likert items. Our data processing approach of treating aggregated Likert-based scores as interval data is supported by established methodological guidelines [27][28][29]. Task completion time, measured as the number of minutes to complete a task, is considered ratio scale data.

We created two datasets, one for novices and one for experts. Each dataset contained participant ID, cognitive load, creativity scores, task completion time, and the type of design tool used (AI or traditional).

Data Analysis:

For both datasets, we employed the Shapiro–Wilk test [30] to assess the distribution of the dependent variables (cognitive load, creativity, and task completion time) in AI and traditional groups. Our rationale for selecting this test is its proven reliability and effectiveness in handling small to moderate sample sizes (less than 50) compared to alternative methods such as the Kolmogorov–Smirnov test [30]. Based on confirmation that all three dependent variables were normally distributed, the independent samples t-test [30] is then selected as an appropriate parametric test to compare the mean scores of the AI and traditional groups.

Results and Discussion:

The following sections present the evaluation and comparison of dependent variables (i.e., cognitive load, creativity, and task completion time) between the two tool groups (AI vs. traditional) within each category (novices and experts), separately. We used IBM SPSS Statistics software [30][31] to conduct all statistical analyses.

Assessment of Data Normality:

We used the Shapiro–Wilk test to assess whether our variables followed a normal distribution. It produces a p-value. If the p-value is less than or equal to the significance level of 0.05, the variable is considered not normally distributed. However, if the p-value is greater than 0.05, the variable is considered normally distributed [30].

Table 2. Summary of normality test (Shapiro–Wilk) for novices

Metric Variable	Tool	Statistic	Sig.
Cognitive load	AI	0.977	0.965
	Traditional	0.902	0.143
Creativity	AI	0.893	0.106
	Traditional	0.946	0.544
Task completion time	AI	0.917	0.226
	Traditional	0.892	0.105

Table 3. Summary of normality test (Shapiro–Wilk) for experts

Metric Variable	Tool	Statistic	Sig.
Cognitive load	AI	0.877	0.148
	Traditional	0.982	0.974
Creativity	AI	0.888	0.190
	Traditional	0.917	0.365
Task completion time	AI	0.891	0.204
	Traditional	0.925	0.438

Tables 2 and 3 show the results of the Shapiro–Wilk test for novices and experts, respectively. In the case of novices (see Table 2), test results show that all three variables i.e. cognitive load, creativity, and task completion time in both AI and traditional groups have p-values greater than 0.05, indicating no significant deviation from normality. Therefore, the normality assumption is satisfied for all metric variables in both groups.

In the case of experts (see Table 3), the test results indicate that all variables (cognitive load, creativity, and task completion time) in both the AI and traditional groups are normally distributed. For instance, within the AI group, cognitive load ($p = 0.148$), creativity ($p = 0.190$), and task completion time ($p = 0.204$) all satisfy the normality criterion ($p > 0.05$).

Assessment of Cognitive Load Using NASA TLX:

For both novices and experts, an independent samples t-test was performed to compare the cognitive load scores of the AI-based UI/UX design tool group and the traditional UI/UX design tool group. Tables 4–7 show the outcomes of this statistical analysis.

Table 4. Group statistics for cognitive load experienced by novices

Metric Variable	Tool	Sample Size	M	SD	Standard Error Mean
Cognitive load	AI	13	2.223	0.525	0.146
	Traditional	13	3.254	0.696	0.193

In the case of novices (see Tables 4 and 5), the results of the group statistics indicate that the AI group reported lower cognitive load values (mean (M) = 2.223, standard deviation (SD) = 0.525) compared to the traditional group (M = 3.254, SD = 0.696). Table 5 shows a p-value of 0.206, which exceeds the 5% significance threshold. This result suggests that the test is not significant, and therefore, the null hypothesis of equal variances is retained, indicating variance equality between the groups. The analysis of equal variances assumed revealed that the mean difference in cognitive load between the AI and traditional tool groups was statistically significant, t(24) = -4.264, p = 0.00027, with a 95% confidence interval (CI) of [-1.52967, -0.53187]. Consequently, the null hypothesis was rejected. Therefore, it can be concluded that the average cognitive load experienced by novices in the traditional group was significantly higher compared to the AI group.

Table 5. Independent samples test for cognitive load experienced by novices

Metric Variable	Variance Equality	F	Sig (p-value)	t	df	2-tailed p-value	Difference of Mean	Std. Error Difference	95% CI [Lower, Upper]
Cognitive load	Equal variances assumed	1.686	0.206	-4.264	24	0.00027	-1.031	0.242	[-1.530, -0.532]
	Equal variances not assumed			-4.264	22.309	0.00031	-1.031	0.242	[-1.532, -0.530]

Table 6. Group statistics for cognitive load experienced by experts

Metric Variable	Tool	Sample Size	M	SD	Standard Error Mean
Cognitive load	AI	9	2.422	0.640	0.213
	Traditional	9	3.789	0.518	0.173

Table 7. Independent samples test for cognitive load experienced by experts

Metric Variable	Variance Equality	F	Sig (p-value)	t	df	2-tailed p-value	Difference of Mean	Std. Error Difference	95% CI [Lower, Upper]
Cognitive load	Equal variances assumed	1.621	0.221	-4.979	16	0.000137	-1.367	0.275	[-1.949, -0.785]
	Equal variances not assumed			-4.979	15.34	0.000154	-1.367	0.275	[-1.951, -0.783]

In the case of experts (see Tables 6 and 7), the observations are similar to those for novices, as the AI group reported lower cognitive load (M = 2.422) compared to the traditional group (M = 3.789). The results of Table 7 show that the p-value (0.221) is greater than the significance level (0.05). Therefore, we fail to reject the null hypothesis, indicating that variance equality exists between the groups. Further analysis (t(16) = -4.979, p = 0.000137, mean difference = -1.367, 95% CI of [-1.949 to -0.785]) revealed that the mean difference in cognitive load between the AI and traditional tool groups was statistically significant.

Figure 2 presents a bar chart comparing the mean cognitive load scores of novices and experts when using a generative AI tool versus a traditional tool. The blue bars represent average cognitive load scores for the AI tool, while the orange bars represent scores for the traditional tool. The chart clearly demonstrates that both novices and experts experienced reduced mental effort when using the AI tool compared to the traditional tool. This answers RQ1.

The black vertical lines with dots represent the standard deviation, showing the spread and variability of cognitive load scores. In this graph, the error bars for the AI tool and the traditional tool barely overlap, or don't overlap at all. This strongly suggests that the difference between AI and traditional methods is clear, meaningful, and likely statistically significant. AI truly lowered workload.

Our findings related to cognitive load are consistent with those of [24], who reported that student participants using only the traditional design tool experienced higher cognitive load compared to participants using AI assistance. Thus, this implies that offering AI support during the prototyping stage can help lower the cognitive load experienced by designers.

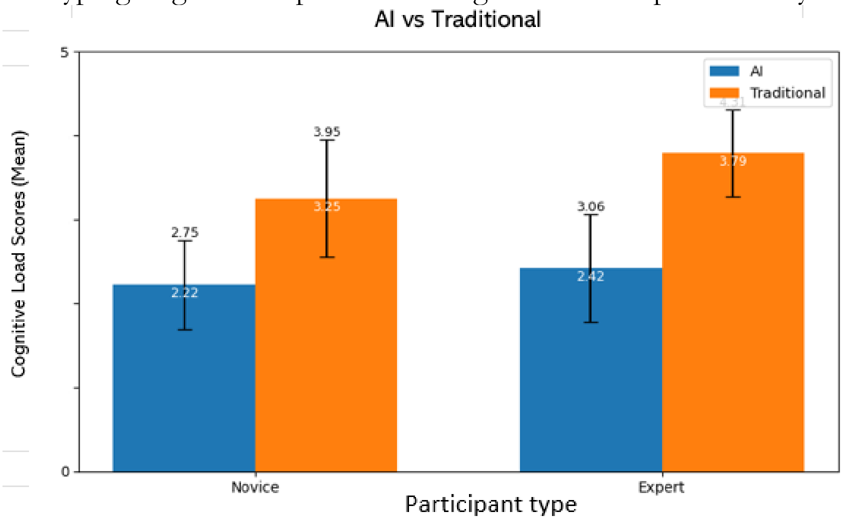

Figure 2. Mean Cognitive Load for Novices and Experts using AI vs. Traditional Tools

Assessment of Creativity Using CPSS:

Tables 8-11 show the results of the independent samples t-tests performed to compare the creativity score between the AI and traditional tool groups for novices and experts. In the case of novices (see Tables 8 and 9), the AI group reported lower creativity (M =3.269, SD = 0.736) compared to the traditional group (M = 3.939, SD = 0.663). The results of Table 8 show that the variance equality exists between the groups since the p-value (0.475) is greater than the significance level of 0.05. Further analysis (t(24) = -2.436, p = 0.023, mean difference = -0.66923, 95% CI of [-1.236 to -.10212]) revealed that the mean difference in creativity between the AI and traditional tool groups was statistically significant.

In the case of experts (see Tables 10 and 11), the observations are similar to those for novices, as the AI group reported lower creativity (M = 3.367, SD = 0.436) compared to the traditional group (M = 4.167, SD = 0.391). Table 10 indicates that the variance between the groups is equal, as the p-value (0.740) exceeds the significance threshold of 0.05. Further analysis (t(16) = -4.101, p = 0.001, mean difference = -0.800, 95% CI of [-1.214 to -0.387]) revealed that the mean difference in creativity between the AI and traditional tool groups was statistically significant. Overall, our findings reveal that both novices and experts produced more creative designs using the traditional tool than with the AI tool, and these differences were statistically significant. This answers RQ2.

Figure 3 presents a bar chart comparing the mean creativity scores of novices and experts when using a generative AI tool versus a traditional tool. The blue bars represent average creativity scores for the AI tool, while the orange bars represent scores for the traditional tool. The chart clearly shows that the traditional tool consistently yields higher creativity scores for both novices and experts. Additionally, experts achieved slightly higher creativity scores overall compared to novices, regardless of the tool used.

The black vertical lines with dots represent the standard deviation, showing the spread and variability of the creativity scores. The overlap between the error bars indicates that although the average scores differ, there is considerable variation in individual participant performance within each group.

Table 8. Group statistics for creativity across design tools for novices

Metric Variable	Tool	Sample Size	M	SD	Standard Error Mean
Creativity	AI	13	3.269	0.736	0.204
	Traditional	13	3.939	0.663	0.184

Table 9. Summary of the independent samples test for creativity across design tools for novices

Metric Variable	Variance Equality	F	Sig (p-value)	t	df	2-tailed p-value	Difference of Mean	Std. Error Difference	95% CI [Lower, Upper]
Creativity	Equal variances assumed	0.527	0.475	-2.436	24	0.023	-0.669	0.275	[-1.236, -0.102]
	Equal variances not assumed			-2.436	23.738	0.023	-0.669	0.275	[-1.237, -0.102]

Table 10. Group statistics for creativity across design tools for experts

Metric Variable	Tool	Sample Size	M	SD	Standard Error Mean
Creativity	AI	9	3.367	0.436	0.145
	Traditional	9	4.167	0.391	0.130

Table 11. Summary of the independent samples test for creativity across design tools for experts

Metric Variable	Variance Equality	F	Sig (p-value)	t	df	2-tailed p-value	Difference of Mean	Std. Error Difference	95% CI [Lower, Upper]
Creativity	Equal variances assumed	0.114	0.740	-4.101	16	0.001	-0.800	0.195	[-1.214, -0.387]
	Equal variances not assumed			-4.101	15.810	0.001	-0.800	0.195	[-1.214, -0.386]

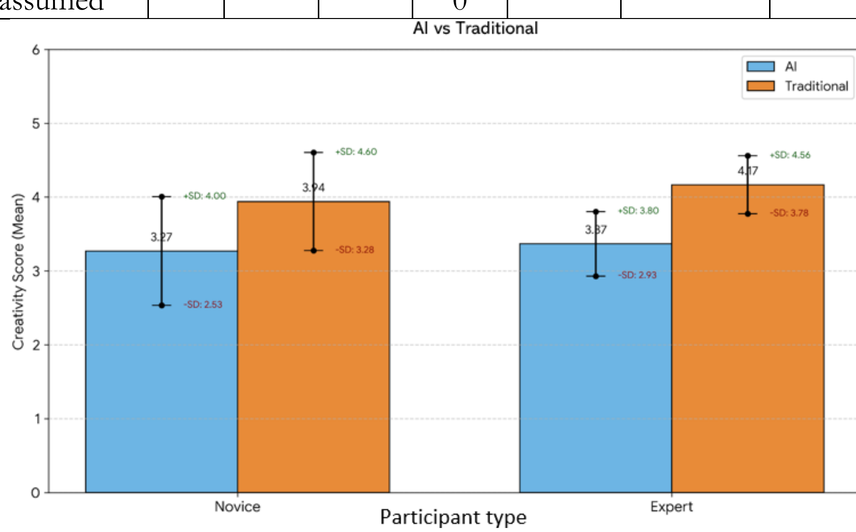

Figure 3. Mean Creativity Scores for Novices and Experts using AI vs. Traditional Tools

Our findings align with [20], who reported that designs developed with the assistance of image search or generative AI tools showed lower originality and variety scores compared to those created entirely without AI support. The reason for lower creativity scores for the AI tool may be the generative AI tool design fixation, as reported by [21]. The repetition or similarities in the design outputs generated by AI tools occur because these tools depend extensively on existing data and learned patterns, which limits their capacity to produce diverse and innovative creative designs.

Assessment of Task Completion Time:

In the case of novices (see Tables 12 and 13), the task completion time for the AI group (24.38) is, on average, lower compared to the traditional group (42.69). The results of Table 13 show that the p-value (0.001) is smaller than the significance threshold of 0.05. Therefore, we reject the null hypothesis, indicating that the variance between the groups is not equal. Further analysis (t(18.006) = -4.456, p = 0.00031, mean difference = -18.308, 95% CI of [-26.940 to -9.675]) revealed that task completion time in the AI group was significantly faster than the traditional group.

Table 12. Group statistics for task completion time across design tools for novices

Metric Variable	Tool	Sample Size	M	SD	Standard Error Mean
Task completion time	AI	13	24.38	6.813	1.890
	Traditional	13	42.69	13.155	3.649

Table 13. Independent samples test for task completion time across design tools for novices

Metric Variable	Variance Equality	F	Sign (p-value)	t	df	2-tailed p-value	Difference of Mean	Std. Error Difference	95% CI [Lower, Upper]
Task completion time	Equal variances assumed	14.828	0.001	-4.456	24	0.00017	-18.308	4.109	[-26.788, -9.827]
	Equal variances not assumed			-4.456	18.006	0.00031	-18.308	4.109	[-26.940, -9.675]

Table 14. Group statistics for task completion time across design tools for experts

Metric Variable	Tool	Sample Size	M	SD	Standard Error Mean
Task completion time	AI	9	23.78	5.563	1.854
	Traditional	9	48.44	7.502	2.501

Table 15. Independent samples test for task completion time across design tools for experts

Metric Variable	Variance Equality	F	Sig (p-value)	t	df	2-tailed p-value	Difference of Mean	Std. Error Difference	95% CI [Lower, Upper]
Task completion time	Equal variances assumed	1.003	0.331	-7.924	16	.000	-24.667	3.113	[-31.266, -18.067]
	Equal variances not assumed			-7.924	14.755	.000	-24.667	3.113	[-31.312, -18.022]

In the case of experts (see Tables 14 and 15), the task completion time for the AI group (23.78) is, on average, lower compared to the traditional group (48.44). The results of Table 15 show that the p-value (0.331) is greater than the significance threshold of 0.05. Therefore, we accept the null hypothesis, indicating that the variance between the groups is equal. Further analysis (t(16) = -7.924, p < 0.001, mean difference = -24.667, 95% CI of [-31.266 to -18.067]) revealed that the mean difference in task completion time between the AI and traditional tool groups was statistically significant. It also revealed that task completion time in the AI group was significantly faster than the traditional group. Overall, our findings revealed that both novices and experts produced designs significantly faster using the AI tool than using the traditional tool. This answers RQ3.

Figure 4 presents a bar chart comparing the mean task completion time scores of novices and experts when using a generative AI tool versus a traditional tool. The blue bars represent average task completion time scores for the AI tool, while the orange bars represent scores for the traditional tool. The chart clearly demonstrates that both novices and experts drastically cut down their task completion times when switching from traditional tools to generative AI.

Interestingly, under the traditional tool, experts took longer on average than novices (48.44 vs 42.69), likely because they followed more detailed or thorough workflows. In contrast, when using the AI tool, both experts and novices completed tasks at nearly the same speed (around 24 minutes), effectively eliminating the performance gap in completion time.

The black vertical lines with dots represent the standard deviation, showing the spread and variability of task completion time scores. In this graph, the error bars for the AI tool and the traditional tool barely overlap, or don't overlap at all. This strongly suggests that the difference between AI and traditional methods is clear, meaningful, and likely statistically significant. AI truly lowered time.

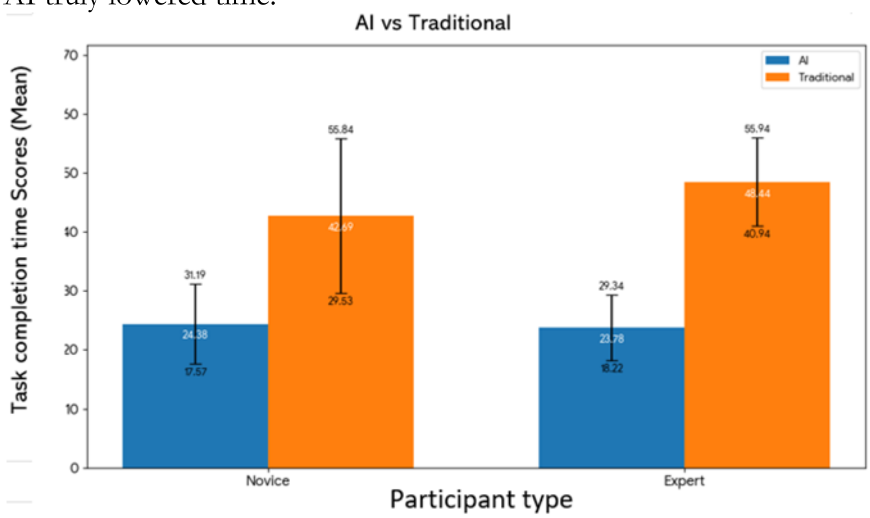

Figure 4. Mean Task Completion Time Scores for Novices and Experts using AI vs. Traditional Tools

Threats to Validity:

We have provided results for both novice and experienced designers who were selected based on their academic or professional background in UI/UX design. Senior students (novice designers) were chosen specifically to avoid issues with tool unfamiliarity, as they had prior experience using the traditional tool Figma. Similarly, experts (experienced designers) had real-world design experience with Figma. Also, both types of designers had earlier experience with generative AI tools for creating UI/UX designs other than Galileo AI.

Participants were randomly divided into AI and traditional tool groups to minimize the sampling bias. We only compared one generative AI tool, Galileo AI, which means our findings may not reflect the capabilities of the wide variety of generative AI tools used in UI/UX design. Nevertheless, we provided results for both novice and experienced designers based on parametric testing, which provide less biased and more precise results.

Implications of the Study:

The findings of this study provide important theoretical and industrial implications for using generative AI tools in the UI/UX design process. Our findings demonstrate that the generative AI tool can significantly improve design efficiency by reducing cognitive load and minimizing the time required to complete UI/UX design tasks. This implies that AI-powered tools can support designers in rapid prototyping, idea generation, and workflow automation, particularly in fast-paced software development environments.

This study has provided empirical evidence that the traditional design approach still outperforms the generative AI tool in terms of creativity. This implies that human designers/practitioners continue to play a critical role in creating engaging UI/UX designs. Furthermore, expert designers' involvement is essential for refining design quality, innovation, and user experience. From the theoretical/academic perspective, this study contributes to the growing body of research on human–AI collaboration in creative work.

Recommendations:

We recommend that practitioners use AI tools such as Galileo AI to accelerate the initial stages of UI/UX design, including wireframing, layout generation, and rapid prototyping, rather than as complete replacements for human creativity. While AI can significantly enhance efficiency and streamline repetitive tasks, human designers and traditional design tools such as Figma should continue to be involved in tasks requiring higher levels of creativity, originality, and emotional engagement. Furthermore, training programs should be introduced to help designers effectively integrate generative AI tools into existing UI/UX workflows.

We recommend that future researchers quantitatively investigate additional AI-powered design tools beyond Galileo AI and Figma to provide broader insights into the effectiveness of AI-assisted design systems. They may also investigate the impact of generative AI tools on other aspects of UI/UX design, such as usability, accessibility, user satisfaction, and emotional engagement.

Conclusions and Future Work:

This study investigated the extent to which a generative AI tool influences the cognitive load, creativity, and task completion time of UI/UX design outputs compared to a traditional design tool, in the context of both novice and expert UI/UX designers. The results of this comparative analysis showed that the generative AI tool significantly reduced both the cognitive load of designers and the time required to create UI/UX designs for both novice and expert participants. For instance, independent samples t-tests showed that the generative AI tool produced lower cognitive load scores than the traditional tool for both novices (2.22 ± 0.53 vs. 3.25 ± 0.70) and experts (2.42 ± 0.64 vs. 3.79 ± 0.52). Similarly, independent samples t-test results indicated that the AI tool reduced average task completion time for both novices (24.38 vs. 42.69) and experts (23.78 vs. 48.44) compared to the traditional tool.

On the other hand, the traditional tool consistently outperformed the generative AI tool in terms of creativity and innovation. For instance, independent samples t-tests results showed that the traditional design tool yielded higher creativity scores than the generative AI tool for both novices (3.94 ± 0.66 vs. 3.27 ± 0.74) and experts (4.17 ± 0.39 vs. 3.37 ± 0.44).

Overall, the AI tool excels in enhancing efficiency; however, it still faces challenges in replicating the depth of human creativity and emotional understanding that are essential for creating engaging UI/UX designs.

In the future, we may examine a wider range of generative AI tools since different tools may produce varying outcomes in terms of creativity, efficiency, and overall UI/UX design quality. We may also perform a comparative study involving multiple generative AI tools and a larger, more diverse participant group to improve the generalizability of the results. The inclusion of participants with varying levels of experience, design expertise, and cognitive approaches may also shed further light on the influence of generative AI tools on UI/UX design processes and outcomes.

Acknowledgement: The authors are thankful to the editor and reviewers for their detailed and useful comments.

Author's Contribution: Iqra Shahzad: Conceptualization, Conducting Experiment, Data Curation, and Writing - Original Draft.

Dr. Marriam Daud: Methodology, Formal Statistical Analysis, Writing - Review & Editing, Supervision.

Wajiha Mughal: Data Preparation, Formal Statistical analysis.

Conflict of interest: The authors have no conflict of interest for publishing this manuscript in IJIST.

References:

[1] H. R. Hartson and P. S. Pyla, "The UX book : Agile UX design for a quality user experience," 2019.

[2] "Figma: The Collaborative Interface Design Tool." Accessed: May 19, 2026. [Online]. Available: https://www.figma.com/

[3] "Adobe XD Learn & Support." Accessed: May 19, 2026. [Online]. Available: https://helpx.adobe.com/support/xd.html

[4] "Sketch · Design, prototype, collaborate and handoff." Accessed: May 19, 2026. [Online]. Available: https://www.sketch.com/

[5] "Midjourney." Accessed: May 19, 2026. [Online]. Available: https://www.midjourney.com/home

[6] "DALL·E 2 | OpenAI." Accessed: May 19, 2026. [Online]. Available: https://openai.com/index/dall-e-2/

[7] "Stable Diffusion Online: The Ultimate AI Image, Video & Editing Platform | Stable Diffusion Online." Accessed: May 19, 2026. [Online]. Available: https://stablediffusionweb.com/

[8] "Stitch - Design with AI." Accessed: May 19, 2026. [Online]. Available: https://stitch.withgoogle.com/

[9] M. A. Runco, "Editorial: Creativity Research: Originality, Utility, and Integration," *Creat. Res. J.*, vol. 1, no. 1, pp. 1–7, Dec. 1988, doi: 10.1080/10400418809534283.

[10] M. A. Runco and G. J. Jaeger, "The Standard Definition of Creativity," *Creat. Res. J.*, vol. 24, no. 1, pp. 92–96, Jan. 2012, doi: 10.1080/10400419.2012.650092.

[11] J. Sweller, "Cognitive Load During Problem Solving: Effects on Learning," *Cogn. Sci.*, vol. 12, no. 2, pp. 257–285, Apr. 1988, doi: 10.1207/S15516709COG1202_4.

[12] R. Heigl, "Generative artificial intelligence in creative contexts: a systematic review and future research agenda," *Manag. Rev. Q.*, vol. 76, pp. 955–992, 2026, [Online]. Available: https://link.springer.com/article/10.1007/s11301-025-00494-9

[13] Y. Luo, "Designing With AI: A Systematic Literature Review on the Use, Development, and Perception of AI-Enabled UX Design Tools," *Adv. Human-Computer Interact.*, 2025, [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1155/ahci/3869207

[14] L. Abrusci, K. Dabaghi, S. D'Urso, and F. Sciarrone, "AI4Design: A generative AI-based system to improve creativity in design–A field evaluation," *Comput. Educ. Artif. Intell.*, vol. 8, p. 100401, 2025, doi: https://doi.org/10.1016/j.caeai.2025.100401.

[15] F. Buendia-Garcia, "Using Generative AI Tools in Collaborative UX design Courses," *Int. J. Artif. Intell. Educ.*, vol. 35, 2025, [Online]. Available: https://link.springer.com/article/10.1007/s40593-025-00518-1

[16] W. F. Yu, "AI as a co-creator and a design material: Transforming the design process," *Des. Stud.*, 2025, doi: 10.1016/j.destud.2025.101303.

[17] M. A. Shahab, "The impact of generative AI on UI design for mobile applications in terms of efficiency using SPACE framework: A survey-based study from mobile application designers/developers," Master of Science Thesis, *Faculty of Information Technology and Communication Sciences, Tampere University,* Nov. 2024.

[18] "(PDF) Concerns and Challenges of AI Tools in the UI/UX Design Process: A Cross-Sectional Survey." Accessed: May 19, 2026. [Online]. Available: https://www.researchgate.net/publication/380522812_Concerns_and_Challenges_of_AI_Tools_in_the_UIUX_Design_Process_A_Cross-Sectional_Survey

[19] J. Li, H. Cao, L. Lin, Y. Hou, R. Zhu, and A. El Ali, "User Experience Design

Professionals' Perceptions of Generative Artificial Intelligence," *Conf. Hum. Factors Comput. Syst. - Proc.*, pp. 1–18, 2024, [Online]. Available: https://dl.acm.org/doi/10.1145/3613904.3642114
[20] S. Wadinambiarachchi, R. M. Kelly, S. Pareek, Q. Zhou, and E. Velloso, "The Effects of Generative AI on Design Fixation and Divergent Thinking," *Conf. Hum. Factors Comput. Syst.*, pp. 1–18, 2024, [Online]. Available: https://dl.acm.org/doi/full/10.1145/3613904.3642919
[21] L. Chen, Y. Song, C. Zheng, Q. Jing, P. Hansen, and L. Sun, "Understanding Design Fixation in Generative AI," *arXiv:2502.05870*, 2025, [Online]. Available: https://arxiv.org/abs/2502.05870
[22] "Enhancing web UI/UX design with AI tools : an in-depth analysis - Theseus." Accessed: May 19, 2026. [Online]. Available: https://www.theseus.fi/handle/10024/891768
[23] Y. Wu, Y. Sun, D. Zhou, X. Yuan, and Z. Wei, "Exploring the Mechanisms of Generative AI's Impact on Design Creativity: An Empirical Investigation Based on the 4P's Model," *Int. J. Human–Computer Interact.*, 2026, doi: 10.1080/10447318.2026.2622577.
[24] T. Chandrasekera, Z. Hosseini, and U. Perera, "Can artificial intelligence support creativity in early design processes?," *Int. J. Archit. Comput.*, vol. 23, no. 1, 2024, [Online]. Available: https://journals.sagepub.com/doi/10.1177/14780771241254637
[25] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," *Adv. Psychol.*, vol. 52, pp. 139–183, 1988, doi: https://doi.org/10.1016/S0166-4115(08)62386-9.
[26] S. P. Besemer and D. J. Treffinger, "Analysis of Creative Products: Review and Synthesis," *J. Creat. Behav.*, vol. 15, no. 3, pp. 158–178, Sep. 1981, doi: 10.1002/J.2162-6057.1981.TB00287.X;ISSUE:ISSUE:DOI.
[27] J. Carifio and R. Perla, "Resolving the 50-year debate around using and misusing Likert scales," *Med. Educ.*, vol. 42, no. 12, pp. 1150–1152, Dec. 2008, doi: 10.1111/J.1365-2923.2008.03172.X.
[28] G. Norman, "Likert scales, levels of measurement and the 'laws' of statistics," *Adv. Heal. Sci. Educ.*, vol. 15, no. 5, pp. 625–632, Dec. 2010, doi: 10.1007/S10459-010-9222-Y/METRICS.
[29] G. M. Sullivan and J. A. R. Artino, "Analyzing and Interpreting Data From Likert-Type Scales," *J. Grad. Med. Educ.*, vol. 5, no. 4, p. 541, Jan. 2013, doi: 10.4300/JGME-5-4-18.
[30] E. T. Berkman and S. P. Reise, "A Conceptual Guide to Statistics Using SPSS," *Sage*, 2012, [Online]. Available: https://sk.sagepub.com/book/mono/a-conceptual-guide-to-statistics-using-spss/toc#_
[31] "IBM Solutions." Accessed: May 19, 2026. [Online]. Available: https://www.ibm.com/solutions