DETERMINING EQUIVALENT DIFFICULTY DESIGN TASKS FOR EXPERIMENTAL STUDIES IN ARCHITECTURAL DESIGN EDUCATION

İhsan Erdem ER[1], Feride Pınar ARABACIOĞLU[2], Hasan TASTAN[3]

[1]Yıldız Technical University Faculty of Architecture, ihsanerdemer@gmail.com
[2] Yıldız Technical University Faculty of Architecture, pinara@gmail.com
[3] Yıldız Technical University Faculty of Architecture, hasantastn@gmail.com

Abstract
Comparative experimental studies in architectural design education often investigate the effects of different design environments on design processes; however, such comparisons are methodologically constrained by learning effects that arise when the same task is repeated. While varying the design task can reduce learning effects, it introduces the challenge of maintaining equivalent task difficulty across conditions. This study aims to address this problem by proposing a systematic approach for identifying architectural design problems with comparable perceived difficulty levels. A single design task was structured around constant spatial constraints, representational requirements, and evaluation criteria, while the main functional program was varied across five alternatives. The study was conducted with 110 second-year architecture students enrolled in the same design studio during a single academic term. Participants evaluated the relative difficulty of the functional variations using a structured pairwise comparison method based on the Analytic Hierarchy Process, and difficulty weights were calculated after excluding inconsistent responses. The results revealed distinct groupings in perceived task difficulty. Micro-scale cafe and exhibition functions were consistently rated as easier, the wildlife observation function was rated as medium, while meditation and resting area functions were perceived as more difficult. These differences were interpreted in relation to functional familiarity and the degree of programmatic definition within architectural education, with more clearly defined and commonly practiced functions associated with lower perceived difficulty. The findings indicate that the perceived difficulty of architectural design tasks can be systematically assessed and controlled, offering a practical methodological framework for selecting equivalent design problems in experimental studies and supporting more reliable comparisons of design environments in future research.

Key Words: *Architectural Design; Task Difficulty; Analytic Hierarchy Process (AHP); Design Task Difficulty Evaluation; Architectural Design Problems*

1. Introduction

Experimental research in architectural design education has been the subject of extensive research aimed at comparing different design environments in order to understand their effects on design processes [1], cognitive demands [2], and learning outcomes [3]. Such studies commonly examine how variations in representation media [4], working conditions [5], or design tools [6] influence designers' performance and experience. However, conducting controlled experiments in architectural design poses distinct methodological challenges, as design tasks are inherently open-ended and sensitive to prior experience, individual interpretation, and learning effects. One of the central methodological problems in comparative design experiments is the learning effect that emerges when participants are exposed to the same design task under multiple conditions. Repeating an identical task may lead to increased familiarity with the problem, solution strategies, or imposed constraints, thereby contaminating the results and weakening causal interpretations. A common strategy to address this issue is to vary the design task across experimental conditions [3], [6], [7] Yet, this approach introduces a second critical challenge: ensuring that the alternative tasks are equivalent in terms of difficulty. If task difficulty is not adequately controlled, observed differences may reflect variations in task demands rather than the effects of the experimental condition itself.

While this issue has been widely recognized in design research, particularly through studies emphasizing the ill-defined and evolving nature of design problems [8] and the co-evolution of problem and solution spaces [9], architectural design research has given relatively limited attention to the systematic identification or construction of design tasks with comparable difficulty levels. In many studies, task equivalence is assumed rather than empirically tested, or it is justified through expert judgment without validation from the target participant group. This situation may limit the methodological robustness and comparability of findings in

experimental architectural design research. The present study addresses this methodological concern by proposing a structured approach to determining equivalent difficulty design tasks for use in experimental studies. The approach is based on maintaining a consistent overall design task structure while varying the functional program and empirically assessing perceived task difficulty through participant-based evaluation. By focusing on perceived difficulty rather than objective performance outcomes, the study aims to capture how design tasks are experienced by students within an educational context.

Accordingly, the main aim of this study is to identify architectural design task variations with comparable perceived difficulty levels that can be used interchangeably in experimental research. The study is guided by the following research question: How do different functional programs within a controlled architectural design task differ in terms of perceived difficulty, and which task variations can be considered equivalent for experimental use? To answer this question, a sketch-based design task with multiple functional variations was administered, and perceived difficulty was evaluated using a structured pairwise comparison method.

2. Method

The study employed a task-based evaluation approach to identify architectural design problems with comparable perceived difficulty levels. A sketch-based architectural design task was developed with fixed spatial constraints, representational requirements, and evaluation criteria to control the overall task structure. The task (Figure 1) required participants to design a transitional space connecting two points with 6-meter level difference within a limited floor area. To introduce controlled variation while maintaining task consistency, the main functional program was varied across five alternatives, including a resting area, meditation space, wildlife (bird) observation point, small exhibition (video and photography), and small cafe.

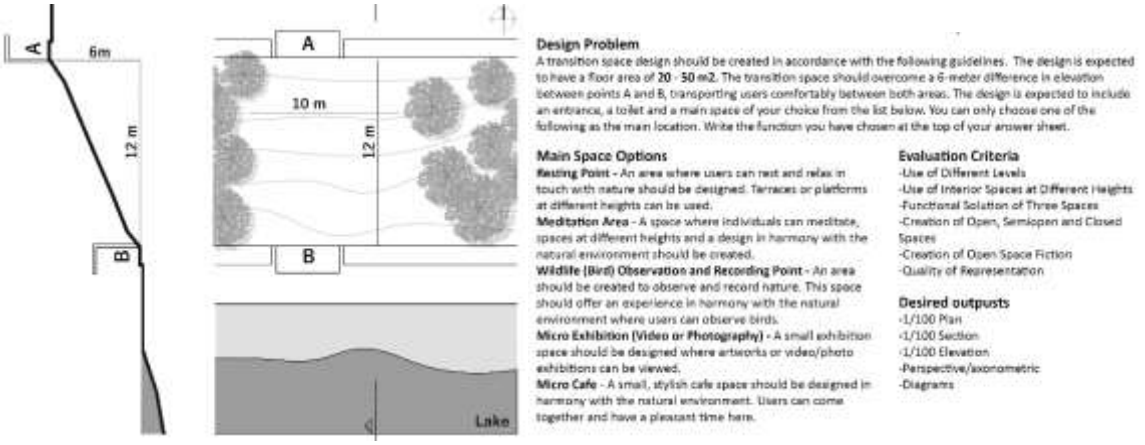


Figure 1: Design task problem

The study was conducted with 270 second-year architecture students enrolled in the same architectural design studio during a single academic term. All participants completed the same design task (Figure 1) under identical conditions as part of an in-class sketch examination. Following the examination, participants were invited to evaluate the relative difficulty of the five functional variations. A structured pairwise comparison questionnaire (Table 1) based on the Analytic Hierarchy Process (AHP) was administered via online survey to assess perceived task difficulty. The post-examination questionnaire was conducted on a voluntary basis, and participants were verbally informed about the purpose of the study and the use of the collected data. No personal or identifiable information was collected.

A total of 110 participants completed the questionnaire, and the AHP-based difficulty assessment and consistency analysis were conducted using these responses. The AHP approach was adopted to reduce the likelihood of random or inattentive responses that may occur in post-examination surveys, as it requires respondents to make consistent relative judgments rather than provide isolated difficulty ratings. In addition, the consistency analysis inherent to AHP allowed the identification and exclusion of responses that did not meet acceptable coherence thresholds. This approach enabled systematic comparison of functional variations by requiring participants to judge relative difficulty rather than assign absolute scores.

Table 1: A structured pairwise comparison questionnaire.

Function	Difficulty									Function
	9	7	5	3	1	3	5	7	9	
Resting Point										Meditation Area
Resting Point										Wild-Life Observation
Resting Point										Exhibition
Resting Point										Cafe
Meditation Area										Wild-Life Observation
Meditation Area										Exhibition
Meditation Area										Cafe
Wild-Life Observation										Exhibition
Wild-Life Observation										Cafe
Exhibition										Cafe

To reduce the influence of random or inattentive responses, consistency ratios were calculated for each questionnaire, and responses exceeding the acceptable consistency threshold of 0.20 were excluded from further analysis. An acceptable consistency threshold of 0.20 was adopted, considering the non-expert participant group and the post-examination data collection context [10]. Following this filtering process, responses from 50 participants were excluded, and difficulty weights were computed based on the remaining 60 participants. These weights enabled the identification of functional variations with similar perceived difficulty levels. The resulting difficulty rankings were used to determine which design task variations could be considered equivalent for use in experimental studies.

3. Results

As part of the study, sketch-based design outputs produced during the examination were collected for all functional variations. Representative examples of these design outputs are presented in Figure 2. However, the primary focus of the present study is not the evaluation or comparison of design quality or performance outcomes. Instead, the collected design outputs serve to document the implementation of the design task and ensure that all functional variations were addressed under comparable conditions. For this reason, detailed analysis of the design outcomes is beyond the scope of this study and is not reported in the results section.



Figure 2: Representative examples of design outputs for each functional variation.

The AHP-based perceived design task difficulty results, calculated using the responses of the 60 participants remaining after the application of the acceptable consistency threshold of 0.20, are presented in Table 2. The results reveal distinguishable differences in perceived task difficulty across the five functional variations. Rather than forming a continuous linear scale, the findings indicate that the functional programs are distributed across distinct difficulty bands based on their relative difficulty levels.

Table 2: AHP-based perceived design task difficulty order (1 indicate hardest, 5 indicate easiest).

Functions	Weight (w_i)	Difficulty order
Resting Point	0.4315	1
Meditation Area	0.2615	2
Wild-Life Observation	0.1462	3

Exhibition	0.1024	4
Cafe	0.0584	5

The resting area function was perceived as the most difficult task variation, followed by the meditation function, which also exhibited a high level of perceived difficulty. Although both functions are positioned at the higher end of the difficulty spectrum, the resting area was rated as substantially more demanding than meditation, indicating that these two functions cannot be considered equivalent in terms of perceived difficulty.

The wildlife observation function occupied an intermediate position within the difficulty distribution. Its difficulty weight was notably lower than those of the resting and meditation functions, yet higher than those associated with the micro-scale exhibition and cafe functions, suggesting a moderate level of perceived difficulty. At the lower end of the difficulty spectrum, the micro-scale exhibition and micro-scale cafe functions were rated as the least difficult task variations. While these two functions exhibited different difficulty weights, they were positioned within the same lower difficulty band when compared to the remaining functional variations. This indicates that, although not identical, these tasks may be considered relatively comparable in terms of perceived difficulty within an experimental context.

Overall, the results suggest that architectural design tasks with varying functional programs do not form a continuous linear scale of difficulty, but instead cluster into broader difficulty bands. Based on the obtained difficulty weights, cafe and exhibition functions were identified as equivalent low-difficulty tasks, meditation and resting area functions as equivalent high-difficulty tasks, and the wildlife observation function as a medium-difficulty task. These groupings provide a practical basis for selecting design task variations with comparable levels of perceived difficulty, rather than assuming strict equivalence, for use in future experimental studies.

4. Discussion

The findings of this study demonstrate the difficulty of achieving strict task equivalence in architectural design experiments and emphasize the importance of empirically assessing perceived task difficulty rather than assuming equivalence. Due to the open-ended and interpretive nature of architectural design tasks, design problems cannot be treated as fixed or stable entities, but are continuously constructed and reframed by designers throughout the design process [9]. As problem and solution spaces co-evolve, the same design brief may be understood, reduced, expanded, or reinterpreted differently by participants, leading to variations in cognitive and experiential demands even when spatial constraints and representational requirements are formally controlled. From this perspective, architectural design tasks differ fundamentally from repeatable problem-solving activities, as design is a constructive activity concerned with how things ought to be rather than how they are and does not aim for reproducibility in outcomes [8]. Consistent with these theoretical accounts, the AHP-based evaluation in this study indicates that variations in functional programs do not produce strictly identical difficulty levels, but instead result in distinct difficulty bands rather than identical difficulty levels.

The observed distribution of perceived difficulty can be interpreted in relation to programmatic definition and familiarity within architectural education. Functions such as micro-scale cafe and exhibition are commonly addressed in studio curricula and are associated with relatively explicit functional requirements and spatial conventions. This familiarity may reduce interpretive ambiguity and intrinsic cognitive load, leading to lower perceived difficulty. In contrast, functions such as resting and meditation spaces require a higher degree of conceptual interpretation, experiential reasoning, and atmospheric decision-making, which are less explicitly defined in conventional functional programs.

These patterns align with design cognition research [11] showing that the level of constraint and definitional clarity in a brief can shape how students distribute cognitive effort between problem framing and solution development. Experimental findings of that study indicate that introducing an external constraint tends to reduce extended exploration of expected behaviours and encourages a more stable, simultaneous focus on

framing and solving, whereas less constrained conditions may involve greater oscillation between these modes over time. From this perspective, more programmatically defined and familiar functions such as a micro-cafe or micro-exhibition may lower perceived difficulty by providing clearer cues for structuring both problem and solution spaces, while more open-ended functions such as resting, and meditation may increase perceived difficulty by requiring greater interpretive work under time-limited sketch conditions.

From a methodological perspective, the findings suggest that architectural design tasks should not be treated as interchangeable without empirical validation of their perceived difficulty. Rather than seeking exact equivalence, the concept of difficulty bands offers a more realistic and applicable framework for experimental design in architectural research. Tasks located within the same difficulty band may be considered comparable for the purpose of minimizing learning effects while maintaining experimental control, acknowledging the inherent variability of design problems.

Several limitations of the study should be noted. Perceived task difficulty was assessed within a single institutional and curricular context, and the findings may reflect characteristics specific to the participants' educational background. In addition, the study intentionally focused on perceived difficulty rather than objective performance or design quality outcomes, as the latter would introduce additional evaluative dimensions beyond the scope of this methodological investigation. Future research may extend this approach by examining the stability of difficulty bands across different educational levels, institutions, or design media, and by exploring relationships between perceived difficulty, cognitive load, and design performance.

5. Conclusion

This study addressed a common methodological challenge in experimental research within architectural design education: the need to vary design tasks to avoid learning effects while maintaining comparable levels of task difficulty. Rather than assuming equivalence between alternative design problems, the study proposed a structured approach for empirically examining perceived task difficulty through controlled functional variation and participant-based evaluation.

Based on the AHP survey assessment, the findings demonstrated that design tasks with different functional programs do not exhibit strictly equivalent difficulty levels, but instead cluster into distinct difficulty bands. The results showed that functions with more clearly defined and familiar programmatic requirements were perceived as less difficult, while more open-ended and experiential functions were associated with higher perceived difficulty. These findings underline the importance of considering functional familiarity and interpretive openness when designing experimental tasks in architectural design research.

From a methodological standpoint, the study contributes a practical framework for identifying design tasks with comparable perceived difficulty, supporting more reliable task selection in experimental studies. By focusing on difficulty banding rather than strict equivalence, the proposed approach acknowledges the inherent variability of architectural design tasks while offering a realistic strategy for minimizing confounding effects related to task difficulty.

Overall, the study demonstrates that perceived task difficulty can be systematically assessed and incorporated into experimental design decisions. This approach has the potential to strengthen the internal validity of experimental studies in architectural design and provides a basis for future research examining the relationships between task difficulty, cognitive load, and design performance.

Acknowledgements

This paper is derived from the PhD dissertation of the first author conducted at the Department of Architecture, Yıldız Technical University. This study was supported by the Yıldız Technical University Scientific Research Projects Coordination Unit under the Doctoral Research Project program (Project No: FDK-2025-6924). The authors gratefully acknowledge their support. We would like to express our sincere gratitude to the Yıldız Technical University Scientific Research Projects Coordination Unit for their valuable support. The authors declare that there is no conflict of interest regarding the publication of this paper. The data supporting the findings of this study are available from the corresponding author upon reasonable request.

References

[1] L. Mejia-Puig and T. Chandrasekera, "The virtual body in a design exercise: a conceptual framework for embodied cognition," *International Journal of Technology and Design Education 2022 33:5*, vol. 33, no. 5, pp. 1861–1882, Dec. 2022, doi: 10.1007/S10798-022-09793-8.

[2] Q. Yang, J. G. Cruz-Garza, and S. Kalantari, "Brain–computer interfaces as an architectural design tool: Feasibility and usability study," *Autom Constr*, vol. 154, p. 105011, Oct. 2023, doi: 10.1016/J.AUTCON.2023.105011.

[3] A. Mohamed-Ahmed, N. Bonnardel, P. Côté, and S. Tremblay, "Cognitive load management and architectural design outcomes," *International Journal of Design Creativity and Innovation*, vol. 1, no. 3, pp. 160–176, 2013, doi: 10.1080/21650349.2013.797013.

[4] R. Yu and J. Gero, "Exploring Designers' Cognitive Load When Viewing Different Digital Representations of Spaces: A Pilot Study," *Smart Innovation, Systems and Technologies*, vol. 65, pp. 457–467, 2017, doi: 10.1007/978-981-10-3518-0_40.

[5] A. Lesage and T. Dorta, "Two conceptual design too ls and an imersive experience: Beyond the pragmatic-pleasurable split in US," *Circuit Bending, Breaking and Mending - Proceedings of the 16th International Conference on Computer-Aided Architectural Design Research in Asia, CAADRIA 2011*, pp. 291–300, 2011, doi: 10.52842/CONF.CAADRIA.2011.291.

[6] A. M. Ahmed, N. Bonnardel, and P. Côté, "The impact of design media on cognitive load during architectural ideation process," *C and C 2011 - Proceedings of the 8th ACM Conference on Creativity and Cognition*, pp. 375–376, Nov. 2011, doi: 10.1145/2069618.2069703.

[7] A. M. Ahmed, N. Bonnardel, and P. Côté, "The impact of design media on cognitive load during architectural ideation process," *C and C 2011 - Proceedings of the 8th ACM Conference on Creativity and Cognition*, pp. 375–376, Nov. 2011, doi: 10.1145/2069618.2069703;ISSUE:ISSUE:DOI.

[8] N. Cross, "Designerly Ways of Knowing: Design Discipline Versus Design Science," *Design Issues*, vol. 17, no. 3, pp. 49–55, Jul. 2001, doi: 10.1162/074793601750357196.

[9] K. Dorst and N. Cross, "Creativity in the design process: co-evolution of problem–solution," *Des Stud*, vol. 22, no. 5, pp. 425–437, Sep. 2001, doi: 10.1016/S0142-694X(01)00009-6.

[10] T. L. Saaty, "Decision making with the Analytic Hierarchy Process," *Scientia Iranica*, vol. 9, no. 3, pp. 215–229, 2002, doi: 10.1504/IJSSCI.2008.017590.

[11] M. Ashrafganjouei and J. S. Gero, "Exploring the effect of a visual constraint on students' design cognition," *Artificial Intelligence for Engineering Design, Analysis and Manufacturing: AIEDAM*, vol. 35, no. 1, pp. 3–19, Feb. 2021, doi: 10.1017/S0890060420000335.