Doctor of Philosophy Dissertation

A Crowdsourcing Method for Architecture

Towards a Collaborative and Participatory Architectural Design Praxis

Jonathan Dortheimer

A Crowdsourcing Method for Architecture:
Towards a Collaborative and Participatory Architectural Design Praxis

To Zoe and Sophie

Gratitude and Acknowledgments

First, I would like to thank my advisors, Prof. Arch. Eran Neuman and Prof. Tova Milo. Prof. Neuman has supported me for over 15 years, from my 4th year as an architecture student through graduate school and, finally, as my doctoral advisor. His visionary approach to architecture, design, and creativity was essential to this project. Furthermore, by acting and demanding the highest standards, Prof. Milo has taught and disciplined me as a rigorous scientist. I was truly fortunate to have such outstanding advisors.

I also wish to acknowledge the anonymous reviewers for their valuable observations that helped me to considerably improve the quality of the dissertation by highlighting critical theoretical issues. I also thank Dr. Anna Ogarkova, for editing this dissertation and teaching me how to write and express myself better.

I am immensely grateful to all my teachers at Tel Aviv University for their valuable insights and feedback. Special thanks go to Dr. Arch. Talia Margalit, my MA thesis advisor and co-founder of Meirim, for her insightful thoughts and support, and to Dr. Arch. Efrat Liebretal-Blumenfeld, who read substantial parts of my dissertation and encouraged me throughout this academic journey.

I also want to thank all students who participated in Tel Aviv University workshops and contributed their talent and skills to my investigation.

Special thanks go to my colleague and friend, Dr. Arch. Tom Shaked, for his inspiration, endless optimism, and challenging projects.

I am beyond words grateful to my parents: my mother, Birgit, who has always been an example of endless intellectual curiosity and my father, Michael, a self-made man, who always encouraged me to fail until I succeeded. I would like to thank my wife, Danna, and my daughters, Zoe and Sophia, for being in my life and reminding me that there is much more in life than work.

Finally, I wish to acknowledge that Tel Aviv University funded this research. The funding gave me the freedom to concentrate on the dissertation work. I would like to thank the donors of Tel Aviv University for their generosity.

Abstract

Recent advancements in information technologies allow exploring new collaborative design methods. In this context, crowdsourcing emerges as an internet-based method that allows undefined crowds to collaborate in producing new information products. Crowdsourcing is achieved by disassembling a complex cognitive task into micro-tasks and relying on collective intelligence for their solution. Since architecture is also an information industry, architects and scholars have identified the potential of information technologies in collaborative architectural design, highlighting the value of a transparent, community-engaging design process.

However, current crowdsourcing methods remain limited in addressing architectural design challenges involving the collaboration of crowds. Architectural crowdsourcing methods are highly experimental or based on traditional methods like competitions or organizations. These methods fall short of integrating the input of various members in the design product and in facilitating this input in the course of the design process rather than at its end. In this regard, there is great potential in advancing crowdsourcing in architectural design.

In order to develop a new collaborative crowdsourcing method for architecture, the research question was formulated as *"What kind of crowdsourcing workflow and micro-tasks are preferable in architectural design to solve the design requirements, provide higher design quality, and is easier to use according to the participants' and expert architects' opinions?"*.

The research question was answered by a literature review of the state-of-the-art in the field, followed by the articulation of a preliminary crowdsourcing model. The preliminary model was applied, measured, studied, and developed in a series of exploration experiments over two workshops. The experiments included the participation of architecture students and experienced architects who were recruited through a freelance website. Each experiment was analyzed and influenced the subsequent one. This process resulted in incremental improvements to the model, which were then tested in the following experiments.

The dissertation presents a large-scale architectural design crowdsourcing approach, platform, and method that allows collaboration between architects with the involvement of non-architect stakeholders.

The workflow that is standing at the core of the platform suggests a new collaborative design process made of three types of micro-tasks that were developed in the presented research. First, 'design' micro-tasks explore the solution space by producing multiple architectural artifacts. Then, 'selection' micro-tasks are used to highlight

the most successful architectural artifacts for further improvement. Finally, 'review' micro-tasks produce improvement ideas highlighting the shortcomings and problems of the selected artifacts.

We suggest the term 'DSR block,' an acronym of Design, Select, and Review, to describe this process above as a universal building block for design crowdsourcing workflows. The exhibited architectural design process implements four kinds of DSR blocks that output different architectural artifacts, naming 3D sketches, 2D sketches, 3D CAD models, and 2D CAD drawings.

Through the iteration within the different artifact kinds, sketches, and CAD artifacts, the design process facilitates design thinking and design consolidation.

Finally, the design process suggests a new open-source-like collaborative and distributed method for architectural design based on the multiplicity of design ideas. The method supports the involvement of non-architects as a crowd participating in the selection and review of tasks critical in directing the design process. The method's outcomes demonstrate the possibility of creating architecture through crowdsourcing and highlight the potential of a new collaborative design process with implications for participatory design, architectural competitions, design process, collective intelligence, and architectural design tools.

Contents

List of Figures

List of Tables

Introduction 1

With the rise of the Internet, new distributed and collaborative production methods have emerged. These methods have been known for producing highly sophisticated and creative digital products. However, they are feasible only with innovation in information management that facilitates synchronizing, organizing, and managing collaborators' efforts. These new online production methods have dramatically changed the culture and economy.

The present dissertation aims to explore how these new online production methods can be used in architectural design. Specifically, we investigate the application of crowdsourcing methods in the creative architectural design workflow.

The introduction chapter presents two key arguments: first, that architecture is a knowledge-based industry (Section 1.1), and second, that architecture is a collaborative praxis (Section 1.2). A short review of the open-source revolution in software is provided to highlight the potential of new ideas in architectural design and the potential of crowdsourcing (Section 1.3). Next, crowdsourcing is introduced as a novel production method (Section 1.4) and the historical roots of crowdsourcing in architecture as public competitions (Section 1.5). Then, research questions are formulated (Section 1.6), and the structure of the dissertation is presented (Section 1.7). The chapter concludes with a discussion of the significance and contribution of the present thesis to previous research (Section 1.8).

1.1 Architecture as Knowledge-based Practice

The first and central argument in this dissertation is that architecture is not only art but also a knowledge-based industry. Therefore, it is beneficial to examine the typical features of other knowledge-based industries and possibly adapt them to architecture. Architecture is an ancient discipline which has emerged thousands of years ago and has since then undergone a series of transformations and revolutions that transformed it into a knowledge-based industry and art. Vitruvius, a Roman architect, argued that the architects' knowledge should go beyond the knowledge of the design of buildings to include also the knowledge of construction, city planning, technical infrastructure, medicine, climate, and even war machines. Hence, the Vitruvian architect had to be knowledgeable and skillful in many disciplines [1].

[1]: Vitruvius (1874), *The Architecture of Marcus Vitruvius Pollio*

A significant turn in the understanding of the profession of an architect came with the writings of Leon Batista Alberti. The Italian Renaissance architect proposed the radical idea that, instead of being master-builders, architects would only produce a design by creating architectural artifacts[1] and provide them to the construction workers [2]. This idea was very odd at that time since the Vitruvian architect was in charge of the entire construction process. Alberti claimed that architects are artists and should produce architectural artifacts in the same way as composers write music. Similar to composers who hand over their compositions to an orchestra and a conductor to perform, an architect provides architectural artifacts to construction workers.

1: e.g., architectural documents like plans, sections, elevations, and models.

[2]: Carpo (2011), *The Alphabet and the Algorithm*

Through this separation of design and construction — or the plan and the building — Alberti's ideas emphasized that architecture is art. As a result of this separation, architectural artifacts became a critical communication protocol between the architect and the builder [2]. In this context, the notion of the Albertian architect transformed architecture into a knowledge-based industry. Indeed, as argued by Marc Uri Porat, similarly to accountants, lawyers, and programmers, architects are knowledge workers who apply existing knowledge to solve new problems[3].

[3]: Porat (1977), *The Information Economy: Definition and Measurement*

The 3rd Industrial Revolution facilitated by computers and communication technologies of the late 20th century disrupted entire industries. This paradigmatic shift in technologies has changed knowledge-intensive industries, such as medicine, law, software, and architectural design. In the mid-1990s, the "paperless studio" of Columbia University's GSAPP[2] experimented with designing using computers. By now, computers have almost replaced pencil and paper and have become the primary working tool used by architects.

2: Graduate School of Architecture, Planning, and Preservation

The digital revolution has been geared by considerable advances in the new computer science. Given that software engineers have relied on architectural design theories, it is necessary to mention two influential books written by the architect and mathematician Christopher Alexander — namely, *Notes on Synthesis of Form* and *A Pattern Language* — that greatly influenced software design theories [4]. Through highlighting the similarities between architectural design and software engineering as knowledge-based domains, these two books paved the way for applying theory from architecture to computer science.

[4]: Gamma et al. (1994), *Design Patterns: Elements of Reusable Object-Oriented Software*

Computer science, particularly in the field of artificial intelligence (AI), is involved in investigating different approaches to solving problems with computers. An intelligent computer should be able to solve ill-structured problems, just as humans do [5]. However, design is an ill-structured problem that is challenging to be understood by a computer [6]. Furthermore, due to the complexity of design, its lack of required knowledge, or the presence of contradictory information that prevents reaching a single and optimal solution, it is also considered a wicked problem. Accordingly, in order to discover how humans solve design problems, researchers traced and formalized different design processes models [7–10]. The most pertinent models are discussed in Chapter 2. In these approaches, design is seen as a process where information is transmitted between the design process components. Each component could be described using 'input,' 'processing,' and 'output' properties.

[5]: Simon (1969), *The Sciences of the Artificial*

[6]: Simon (1973), "The structure of ill structured problems"

[7]: Takeda et al. (1990), "Modeling design processes"

[8]: Maher et al. (1996), "Formalising Design Exploration as Co-evolution: A Combined Gene Approach"

[9]: Dorst et al. (2001), "Creativity in the design process: Co-evolution of problem-solution"

[10]: Gero et al. (2004), "The situated function-behaviour-structure framework"

Since such models view the tacit activity of designing as an explicit process, they are fundamental for the present study, as these models offer a possibility to computerize and distribute the design process components as micro-tasks in a crowdsourcing workflow. Moreover, the modulation and exchange of information open up new ways to systematically explore the crowdsourcing design process [11].

[11]: Maher (2011), "Design Creativity Research: From the Individual to the Crowd"

Recent decades have witnessed the advances of new technologies, such as Machine Learning (ML), automation, and cognitive computing [12], that have collectively been referred to as the 4th Industrial Revolution. While, at present, the use of AI in architectural design remains very experimental [13, 14], we expect that, in the future, it will become a natural part of the design process. This raises many questions, particularly those related to the relationship between architects, clients, community, and Artificial Intelligence.

[12]: Picon (2016), "From Authorship to Ownership: A Historical Perspective"

[13]: As et al. (2018), "Artificial intelligence in architecture: Generating conceptual design via deep learning"
[14]: Chaillou (2019), "AI + Architecture | Towards a New Approach"

This dissertation deals with crowdsourcing, a computer-managed process that employs collective human intelligence to solve cognitive challenges. The crowdsourcing process enables integrating stakeholders and architects into a design process that is formulated as an algorithm. We suggest that the process's algorithmization is a fundamental step toward developing better artificial intelligence components for a creative architectural process.

1.2 Paradigm Shift in Architecture

The second key argument in this dissertation is that architecture is and has always been characterized by a high level of teamwork and collaboration. Based on this argument, we suggest that a crowdsourcing process for architecture may be a natural stepping stone in the evolution of the discipline into a more digital and collaborative praxis.

Previous research has demonstrated several ways in which architectural work is essentially collaborative. Since architects mostly work to meet their clients' demands, this requires close collaboration with the clients, which results in the establishment of shared ownership of both parties with respect to the created designs [15]. Moreover, creating large and complicated structures involves collaboration in the design process, which results in a division of labor and authorship [16]. A particularly high level of teamwork is required within the design team, where the design work is divided among the architect, engineer, 3D artist, model builder, licensing expert, and draftsperson [17].

[15]: McDonnell (2009), "Collaborative negotiation in design: A study of design conversations between architect and building users"

[16]: Picon (2016), "From Authorship to Ownership: A Historical Perspective"

[17]: Smith et al. (2017), *Leading Collaborative Architectural Practice*

As mentioned above, initially, architects were the master-builders — i.e., experts in all building systems they designed and built from the ground up. However, architects largely depended on artisan and artists and were limited in their creativity by the overall organization of exterior and interior decor [18]. Accordingly, a high level of collaboration among all involved parties was required at that time.

[18]: Picon (2013), *Ornament: The Politics of Architecture and Subjectivity*

With the 2nd Industrial Revolution and technological development of the 19th century, buildings became more complex and required a new level of expertise [17]. Accordingly, new architectural specializations — such as engineers, urban planners, interior designers, among others — emerged. Today, buildings have become even more complex from the technological and regulatory standpoints and require an even higher level of collaboration among multiple experts.

[17]: Smith et al. (2017), *Leading Collaborative Architectural Practice*

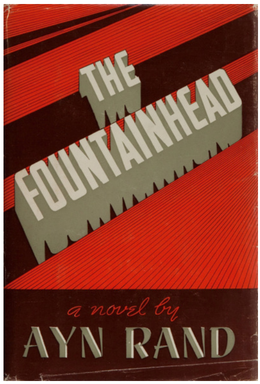


Figure 1.1: Cover of the first edition of the Fountainhead by Ayn Rand (1943).

Following the new requirements for a more complex architectural praxis, the division of architects and artisan and builders became regulated by law, thereby changing how an architect was educated and who could be named an architect [19]. The new laws distinguished between the builder and the architect and set forth the architect's responsibilities and obligations.

The formalization of this shift from master-builders to professional architects created the modern image of the Architect [19]. The best example of this novel image is the imaginary heroic architect Howard Roark, the protagonist of Ayn Rand's "The Fountainhead" (1943). Roark, whose figure was arguably based on the architect Frank Lloyd Wright, represented Rand's ideal of the

[19]: Saint (1983), *The Image Of The Architect*

[19]: Saint (1983), *The Image Of The Architect*

individual: a person who adheres to his/her personal truth without compromises. Architects have traditionally been recognized and praised for a similar idealization of the individual author and the idea of creativity as an individual effort. However, soon afterward, the notion of a genius artist became the target of criticism.

For instance, Christopher Alexander argued that design problems encountered by architects are complex, which makes it almost impossible to solve them without an evolutionary, collaborative, and iterative process[20]. Furthermore, and from a different perspective, Jane Jacobs reasoned that architecture and planning had become disconnected from their community and resulted in non-human architecture [21]. In addition, contemplating upon the beauty and livelihood of vernacular architecture, Bernard Rudofsky went as far as to suggest that professional architects might not be needed[22]. This and other criticisms of the notion of a genius architect had a dramatic effect on the academic and professional discourse about architecture [23].

[20]: Vayssiere et al. (1977), "Notes on the Synthesis of Form"

[21]: Jacobs (1961), *Death and life of great American cities*

[22]: Rudofsky (1965), "Architecture without Architects: A Short Introduction to Non-Pedigreed Architecture"

[23]: Reich et al. (1996), "Varieties and issues of participation and design"

In the subsequent years, designers and architects looked for novel methods to more appropriately designing their projects for future users [24]. In the entirety of these various new methods, the following two major approaches can be identified. The first approach is User-Centered design (UCD), in which experts observe users and scientifically draw conclusions. The second approach is Participatory Design (PD), in which participants are treated as partners. The PD approach is based on various design tools, such as games, cards, models, collages, and so forth, all of which make design tasks more accessible to users [25]. However, participatory design still has many challenges, including the politics of design, tacit knowledge [3], as well as methods, tools, and techniques [26].

[24]: Sanders et al. (2008), "Co-creation and the new landscapes of design"

3: e.g., a knowledge gap that will be discussed later

[26]: Kensing et al. (1988), "Participatory Design: Issues and Concerns"

In summary, despite the widely held view that master-builders of the past were geniuses who created concepts of entire buildings, there was a great deal of collaboration between architects, and artisans, or between professionals and stakeholders. Therefore, architecture has always been and still is the product of (sometimes complex and challenging) cooperation.

1.3 The Open-Source Revolution

In this section, we explain the technological approaches that underlie our proposed architectural crowdsourcing method. The core of this method is the 'Open-Source' approach. The term 'Open-Source' was coined at the end of the 1990s in continuation of the 'free software' movement ideas. The advocates of the 'free software' movement aimed to create alternatives to the legal framework protecting intellectual property (IP) and business models in the software industry based on rigid and exclusive proprietary regulations. The copyright-based business model enabled the sale of software under a restrictive license, which limited the rights of the customer to only the use of that software (but not ownership of that software). The copyright business model does not provide the software's 'source code' [4] which, while guaranteeing the protection of intellectual property, prevents introducing changes, adaptations, and fixes to the software, as well as not allowing users to learn how it works.

4: A human-readable programming language, in contrast to the 'binary machine code' which can be executed, but is not readable by humans

In contrast, the 'free-software' movement offered a new licensing framework that ensured that *free software* is always distributed with its 'source code,' allowing customers to fix, improve, develop and adapt the software for any use [27]. The licensing framework was also accompanied by an open call to establish a free operating system called GNU, better known as Linux.

1.3.1 Open-Source Production Method

In his influential book *The Cathedral and the Bazaar*, Eric Raymond revealed the unique collaborative production process followed by the free software community [28]. The author described the development of the operating system Linux in a decentralized fashion by thousands of programmers from all over the world. Raymond (1999)

[28]: Raymond (1999), *The cathedral and the bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*

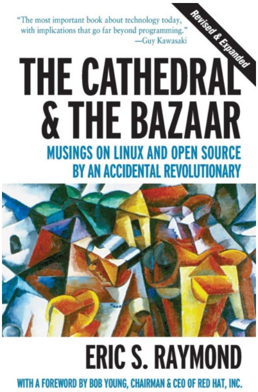


Figure 1.2: Cover of the The Cathedral & the Bazaar by Eric S. Raymond (2001).

also credited Linus Torvalds, known for producing the Linux kernel, for creating this distributed development method. Raymond summarized 19 'lessons' to be learned from the unique production method. The following seven of these lessons are particularly relevant to architecture and PD.

- "*Treating your users as co-developers* is your least-hassle route to rapid code improvement and effective debugging" [28, p. 27].
- "Release [the software] early. Release often. And *listen to your customers*" [28, p. 29].
- "Given a *large enough beta-tester and co-developer base*, almost every problem will be characterized quickly, and the fix obvious to someone" [28, p. 30].
- "If you treat your *beta-testers as if they're your most valuable resource*, they will respond by becoming your most valuable resource" [28, p. 38].
- "The next best thing to having good ideas is *recognizing good ideas from your users*. Sometimes the latter is better" [28, p. 40].
- "Any tool should be useful in the expected way, but a *truly great tool lends itself to uses you never expected*" [28, p. 44].
- "... many heads are inevitably better than one" [28, p. 54].

Raymond's (1999) book catalyzed the growth of interest within the software industry to the progress of the 'free software' movement. However, the radical standpoint of the movement concerning IP hindered commercial use of free software, as its license allowed customers to distribute the software for free to anyone. This standpoint led to the development of middle-ground methods with the declaration of 'Open-Source.' The term 'Open-Source' stands for a libertarian approach regarding the possibility of distributing 'open' software commercially by including licenses designed to allow distribution of source code and encourage users to join the design process, and later 'close' the source code and sell it. Due to the success of these methods, traditional software companies of today frequently join and participate in open-source projects.

1.3.2 The Open-Source Way

There have been many studies that aimed to understand why individuals would be willing to freely share work and talent and, even more interestingly, why commercial companies would share innovations and IP with their competitors. For instance, Eric Von-Hippel and Georg Von-Krogh (2003) proposed the 'private-collective' innovation model that explains the value of sharing technological innovations to inventors [29]. According to this model, open-source provides a framework for sharing knowledge, leading to increased productivity, enabling inventors more free time for innovation, improving the development process, and, finally, the end product. More importantly, Von-Hippel and Von-Krogh (2003) demonstrated that organized sharing of inventions between entrepreneurs is not an innovation of the open-source movement but existed in the past [30].

[29]: Hippel et al. (2003), "Open Source Software and the "Private-Collective" Innovation Model: Issues for Organization Science"

[30]: Hippel (1987), "Cooperation between rivals: Informal know-how trading"

Today, open-source use has become a shortcut for many programmers who have been able to compete and provide high added value at a lower cost to customers. Accordingly, the idea of open-source has encouraged innovation, becoming thus a central feature in the digital revolution. Furthermore, it is presented as a philosophy, 'The Open-source Way,' as a culture and framework for co-operative work that respects both the programmer and the customer. This approach advances transparency and promotes decentralized, fast, and collaborative development[31]. This evolution has led to both theoretical advances and implementations of these ideas in various fields [32], including robotics, electronics, pharmaceuticals, education, design, and architecture. The open-source method has also inspired new models of decentralized production of knowledge, also known as crowdsourcing [33], which will be discussed in the next section.

[31]: Red Hat Inc. (2009), *The open source way*

[32]: Lessig (2004), *Free Culture*

[33]: Howe (2006), "The Rise of Crowdsourcing"

1.3.3 Open-Source Architecture

In 2003, i.e., 18 years after the introduction of 'Free software' and with the emergence of the 'open-source' production models, the Dutch architect Dennis Kaspori proposed adopting the open-source model for architecture [34]. Kaspori (2003) theorized that implementing the open-source architecture would create a "learning organization" that will "offer an alternative model in which innovation is achieved through the active participation of all parties". Although Kaspori's (2003) proposal was compelling, he did not explain how this model could be applied to architecture.

[34]: Kaspori (2003), "A Communism of ideas Towards an open-source architectural practice"

Carlo Ratti and Associates (2011) further enriched the open-source debate. Based on the idea of open-source, they proposed a new "crowd-funded" planning paradigm that relied on amateurs and consumers collaborating [35]. They argued for 'smart' houses based on open software and hardware that could be adjusted to fit the different needs of their successive residents. Their discussion focused on the democratic and decentralized production method of the software development world. Despite their impressive vision, they failed to address any practical application, leaving their vision theoretical and intangible.

[35]: Ratti et al. (2011), "Open Source Architecture"

In their 2015 book "Open Source Architecture," Ratti and Claudel (2015) integrate the idea of open-source architecture with criticism of modernist planning and postmodernist theories of involving users in the design process [36]. They affirm that architects have worked with large economic forces but lost touch with the real needs of the clients and community, arguing that communities have always built houses for themselves without architects. As a result, communities could play a key role in architectural planning. They proposed a theoretical model for open-source architecture that allows users to be included in the design process. According to this 'choral architect' model, end-users will plan buildings, while architects supervise and coordinate as facilitators, not as creators.

[36]: Ratti et al. (2015), *Open Source Architecture*

The theme of Open-source Architecture was also explored in a special issue of Architectural Design in 2016. As digital architectural creations become more advanced and affordable, 3D printing creates opportunities for the duplication of design discussing open source and intellectual property. Garcia (2016) proposed an architectural patent system that would allow innovators to receive micro royalties while promoting innovative design [37]. Aaron Sprecher and Chandler Ahrens (2016) praised open-source technological progress for advancing versatility, dissolving disciplinary boundaries and hierarchies, and offering an alternative to traditional intellectual property models [38]. Wendy Fok (2016) discussed digital fabrication, digital collaborative platforms, and architectural applications of the 'Internet of Things' [39]. Finally, Antoine Picon (2016) stated that authorship questions are already disputed in architecture since 'Starchitects' present their projects as being the work of their talent when they hire hundreds of professional architects to collaborate to produce them [16].

[37]: Garcia (2016), "Architectural Patents and Open-Source Architectures: The Globalisation of Spatial Design Innovations (or Learning from 'E99')"

[38]: Sprecher et al. (2016), "Adaptive Knowledge in Architecture: A Few Notes on the Nature of Transdisciplinarity"

[39]: Fok (2016), "Opening Up the Future of Open Source: From Open Innovation to the Internet of Things for the Built Environment"

[16]: Picon (2016), "From Authorship to Ownership: A Historical Perspective"

In a recent study, Dortheimer and Margalit (2020) reviewed the ways in which open-source production models have been used in architecture. Having examined most of the 'open-source' architecture projects, they found that the 'open-source' model was successful in developing technological innovations [5] [40]. However, the model did not succeed with architectural design. Declaring a project as open source and providing blueprints free of charge had little effect on the outcome and did not generate the expected community that would continue developing the design. For instance, soon after winning the Pritzker Prize, the laureate Alejandro Aravena declared four 'incremental housing' projects as 'open-source'. However, these projects did not show any properties of being truly 'open-source'; in addition, there was no sign of a community that continued to develop and support these designs.

5: for example, WikiHouse.cc, which established a novel fabrication method

[40]: Dortheimer et al. (2020), "Open-source architecture and questions of intellectual property, tacit knowledge, and liability"

Dortheimer and Margalit (2020) suggested that *open sourcing* architecture has limited benefits, first because architecture has never been 'closed,' as architectural knowledge was freely shared since the rise of humanity as suggested by Von Hipple and Von Krough (2003). Second, artistic and moral values like 'Inspiration' and 'Plagiarism' have emerged in the discipline that protects IP, making laws

Figure 1.3: Wikihouse pavillion. Credits: photo by Jan Willem de Groot (CC-BY license)

not essential to promoting innovation in architecture. In summary, while the 'open-source' model in architecture does not facilitate the desired collaborative practice as envisioned by Kaspori (2003), crowdsourcing might produce such design collaboration [40].

1.4 The Crowdsourcing Method

In 2000, Richard Stallman, founder of the free software movement, proposed creating the "Nupedia", an online encyclopedia based on free software principles that would allow any user to contribute to an encyclopedic entry [41]. As "Wikipedia" adopted the license of free software movement, the projects merged [42]. The dynamics in which entries are written and edited in Wikipedia are similar to the dynamics Raymond (1999) described using the bazaar metaphor: some users with particular knowledge contribute their time and talent to improve

[41]: Stallman (2000), *The Free Universal Encyclopedia and Learning Resource*

[42]: Stallman (2014), *The Free Encyclopedia Project*

and expand Wikipedia, while other volunteers monitor the changes to prevent abuse.

Following the success of Wikipedia and open-source software, Jeff Howe coined the term 'crowdsourcing' in his article "The Rise of Crowdsourcing' [33]. This article describes an emerging production method where traditional work performed by an employee is outsourced to a large, distributed and undefined group of people. The new term referred to an existing Internet phenomenon where many sites used their audiences to generate new knowledge, or to the practice of publishing an 'open call' to receive information proposals from the public. Due to multiple interpretations of this term and many applications that use the data derived from the 'crowd,' there is currently no single agreed and explicit definition of 'crowdsourcing' [43].

[33]: Howe (2006), "The Rise of Crowdsourcing"

Of note, crowdsourcing is not considered open-source, as the products it creates are not 'open' or 'free' as defined in the Open Source Definition [44]. Instead, crowdsourcing is an Internet-based production method that allows many people to perform small tasks (called micro-tasks) compiled into a larger product. The term itself is related to outsourcing as a business process where a job is outsourced to a crowd instead of assigned to a subcontractor.

[44]: Perens (2008), "The Open Source Definition"

The main advantage of crowdsourcing is that it makes it possible to reach, recruit, and engage a vast and global human intelligence pool that can propose new solutions and original products [45]. Another advantage of crowdsourcing is cost-effectiveness, which arises from the precise budget allocation and the accuracy of the requested services rendered as micro-tasks [45].

[45]: Milo (2011), "Crowd-Based Data Sourcing"

1.5 Crowdsourcing of Architectural Design

Crowdsourcing is not fundamentally new to architectural design [46]. Public contests, which are based on an 'open call' to an undefined crowd, are a common and accepted

[46]: Angelico et al. (2012), "Crowdsourcing Architecture : a Disruptive Model in Architectural Practice"

practice to generate and evaluate innovative design solutions. Also, architectural competitions are an essential part of the training of architects in higher education institutions [47]. Given that competitions are part of the architects' tradition, several architectural crowdsourcing websites have emerged offering to arrange public online competitions (e.g., Arcbazar, GoPillar, CoContest, and Popularc). However, while the competition model is valuable for architectural innovation, it facilitates neither the collaboration between architects nor the involvement of stakeholders except for defining the requirements and selecting the winner.

[47]: Guilherme (2014), "Competitions serve a larger purpose in architectural knowledge"

Accordingly, the practice of public architecture competitions has received a great deal of criticism. For instance, Magnus Rönn (2009) argued that in architectural competitions, there are mostly multiple good solutions to the design problem, and the challenging task of the jury is to select the best design taking into consideration various conflicting interests. Therefore, the decision process of the jury rarely results in the selection of an overwhelmingly superior design [48]. Likewise, Peggy Deamer (2015) argued that "the winner takes it all" model is wasteful and unfair to the rest of the participants [49]. Furthermore, Elizabeth Keslacy (2018) highlighted ethical issues that arise from the commercial crowdsourcing websites that perform online architecture competitions [50]. The author claimed that the website promises non-existing cultural capital gains and converts highly regulated artistic and creative work to speculative work that, on average, has one-third of return, if any at all.

[48]: Rönn (2009), "Judgment in the Architectural Competition – rules, policies and dilemmas"

[49]: Deamer (2015), "The Guggenheim Helsinki Competition: What Is the Value Proposition?"

[50]: Keslacy (2018), "Arcbazar and the Ethics of Crowdsourcing Architecture"

Therefore, despite the fact that architectural competitions are a driving force for architectural innovation, they are limited and problematic. In this context, there is an urgent need to investigate the potential of crowdsourcing methods in architectural design.

1.6 Research Questions

This dissertation aims to produce an architectural crowdsourcing workflow that overcomes the shortcomings of the competition model. To this end, we review relevant literature, formulate a preliminary model, and develop the formulated model over multiple experiments. The main research questions addressed in the present study are as follows:

Main RQ

What kind of crowdsourcing workflow and micro-tasks are preferable in architectural design to solve the design requirements, provide higher design quality, and is easier to use according to the participants' and expert architects' opinions?

This research question is addressed by conducting multiple experiments (see Chapter 5). Since the experiments focus on the specific aspects of the micro-tasks that are an integral part of crowdsourcing workflows, the following five specific research questions are formulated:

RQ-1

Which type of design micro-task yields artifacts that are evaluated higher by experts?

RQ-2

Which type of selection micro-task yields an artifact selection that is closer to expert evaluation?

RQ-3

Which type of review micro-task yields design reviews that are beneficial to designers?

RQ-4

In which parts of the architectural crowdsourcing workflow do professional participants provide better performance and results as compared to non-professional participants?

RQ-5

What kind of crowdsourcing workflow and micro-tasks are preferable in architectural design to solve the design requirements, provide higher design quality, and is easier to use according to the participants' and expert architects' opinions?

These research questions are addressed by a series of experiments (see Chapter 5).

1.7 Structure

The research process, which gave rise to the structure of the present dissertation, is shown in Figure 1.4.

Chapter 1 provides an introduction to the new open-source production methods and discusses their potential for a new architectural praxis.

Chapter 2 offers a comprehensive literature review. Different crowdsourcing methods are reviewed with an emphasis on creativity. Some crowdsourcing studies in various domains, such as software engineering, content writing, and design, are also presented. The theoretical foundation of the present dissertation is laid down based on architecture design, creativity in design, collaboration in design and architecture, and knowledge theories.

Chapter 3 presents a pilot experiment conducted using 'Arcbazar,' a commercial architecture crowdsourcing website. The main aim of this experiment was to evaluate the research methods and learn about the benefits and shortcomings of the existing crowdsourcing workflow.

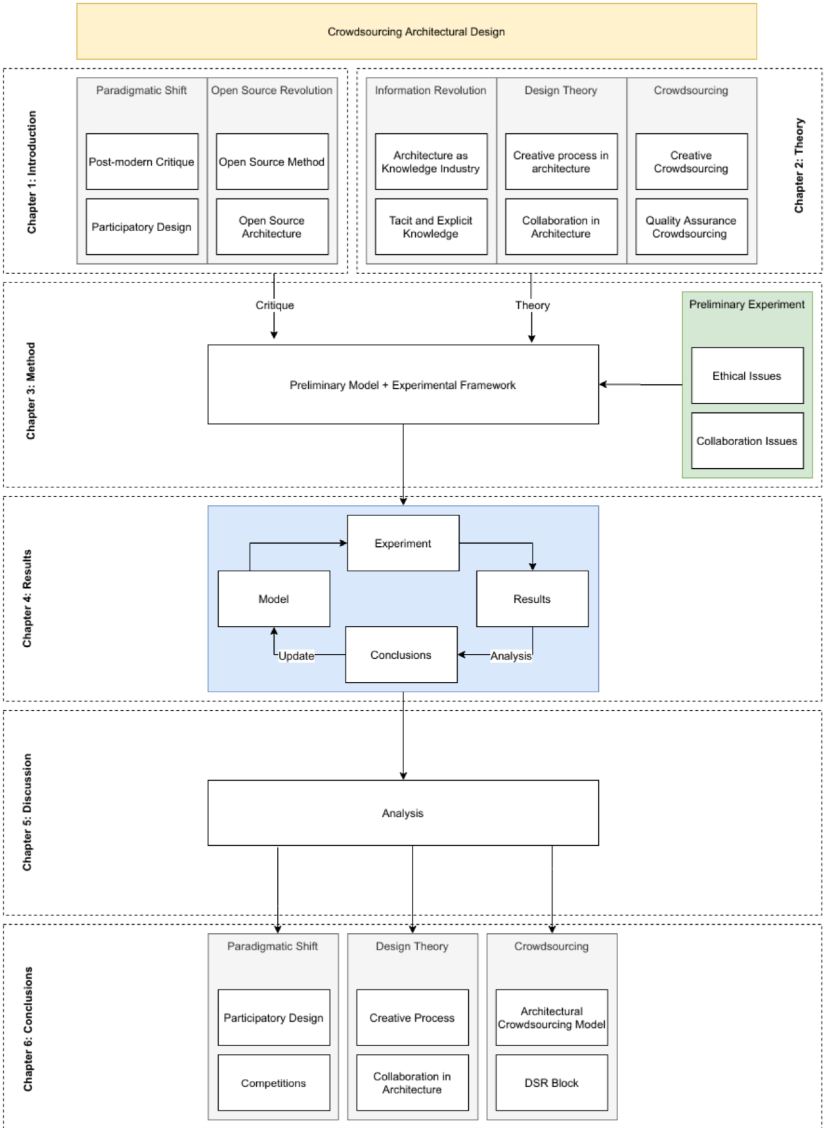


Figure 1.4: Research Structure

Chapter 4 builds upon the literature presented in Chapter 2 and the conclusions drawn from the results of the experiment described in Chapter 3. In this chapter, we formulate a preliminary crowdsourcing model and present the experiments, tools, and methods that were used.

Chapter 5 presents the results of our experiments. The aims, methods, and results of each experiment are provided, and the conclusions are highlighted.

Chapter 6 discusses the implications of our experimental results and outlines a new creative crowdsourcing model specific to architectural design. This model is also discussed in the light of previous work described in the literature review chapter (Chapter 2).

Finally, in Chapter 7, the conclusions are drawn. We highlight some evidence and discuss the lessons about architects' design process, architectural public participation, and collaborative architectural praxis.

1.8 Significance and Contribution

The results of the present dissertation contribute to the available body of work on design crowdsourcing techniques. Our main focus is on the crowdsourcing of architectural design as a design process.

Specifically, in the present dissertation, the design process is articulated as an algorithm consisting of micro-tasks. We define a framework for designing design processes in crowdsourcing called the DSR block. This framework is generic and could be used for other design methods.

We formulate and characterize design micro-tasks in terms of user interfaces and information provided as input. Considering time limitations, a micro-task simply and effectively explains the requirements to a user (a designer or architect). The design of such micro-tasks is essential for the success of complex design processes.

This study also addresses the integration of professionals and non-professional stakeholders in the process of creating a sophisticated architectural design. The proposed method offers significant participation through the crowdsourcing technique in navigating the development process and providing feedback to designers.

It is widely known that sketching supports design thinking. In this respect, the present study contributes to the knowledge of design techniques by offering a computer-based process that alternates between analog sketching and digital representation.

Furthermore, the proposed design process allows for the decentralization of the design process, thus enabling the participation in the design process of many agents through the Internet without creating any interdependence among the participants. This aspect of the proposed model is particularly beneficial for the creation of new online workflows for geographically distributed teams and individuals.

Next, the results of the present dissertation contribute to the knowledge of Participatory Design by offering a new digital technique that stakeholders can employ without the need to engage in complex learning activities and without dealing with the politics of design. We redefine the concept of "open-source architecture" as an approach that leverages information technologies to empower designers' and users' active participation in the shared and collaborative human-centered design process, fostering collective intelligence.

Finally, the design process proposed in the present dissertation is introduced as an algorithm. We suggest that the algorithmization of the architectural process is an initial but important step in understanding the design process towards artificial intelligence in architectural design.

Theory | 2

This chapter outlines the theories and concepts associated with the architectural crowdsourcing process proposed in the present study.

The discussion starts with several design-process theories essential to the development of a crowdsourcing process (Section 2.1), with a particular focus on design models based on collaboration and teamwork, as well as some architecture-specific models. Quality evaluation and critique are highlighted as essential parts of the process.

Theories and research in architecture that discuss multiple kinds of collaboration are reviewed (Section 2.2). Based on this discussion, we conclude that architecture is almost always the product of several types of collaboration, which suggests the potential for developing a new collaborative method.

Next, we review relevant knowledge theories based on Michael Polanyi's *Tacit Knowledge* theory (Section 2.3). Given the idea that crowdsourcing may be used for public participation, the here presented theories clarify and explain the differences between professionals and stakeholders and support the participants' different roles in the design process.

Finally, Section 2.4 introduces several creative crowdsourcing methods that include complex activities such as article writing, software development, and problem-solving. We also review several crowdsourcing quality evaluation methods and methods of providing feedback, both of which are important parts of the design process.

2.1 The Design Process

In order to create a new crowdsourcing production model for architecture, we need to thoroughly investigate the design process, seen in the present thesis as a major generator of architectural artifacts. The design process, which incorporates various systemic design methods to produce design, has been actively investigated since the 1960s [51].

[51]: Cross (1977), *The Automated Architect*

Following the Design Methods conference held in 1962, there has been a broad consensus among scholars that a systemic design process consists of the following three stages [51]. In the first stage, the analysis of the problem at stake is made. In the second stage, solutions are synthesized. Finally, in the third stage, an evaluation is performed to select the best solution [52].

[52]: Page (1963), "A review of the papers presented at the conference"

For instance, the *method of systemic design* proposed by Jones (1963) consisted of the three stages as mentioned earlier of analysis, synthesis, and evaluation [53]. Jones' (1963) method was an attempt to unify traditional and rigorous mathematical approaches. More specifically, this method allows the designer to focus on solving problems by providing a system of notations that records every item of design information outside of the memory. The first step (analysis) includes collecting, classifying, and mapping the relationships between factors, articulating the problem specifications, and reaching an agreement. In the second step (synthesis), creative thinking is applied to perform partial solutions considering limitations. Finally, in the third step (evaluation), the solution is judged by various evaluation methods.

[53]: Jones (1963), "A method of systemic design"

However, upon testing methods like the one proposed by Jones (1963), several practitioners argued that linear systemic methods were too formal and impractical [51]. Specifically, a series of observational studies established that, in practice, the design process has interdependency links between decisions, which require a re-evaluation of previous decisions once design element changes are implemented [54]. This interlinked structure of the design process suggests that the process is cyclical and iterative (see Figure 2.1).

[54]: Luckman (1967), "An Approach to the Management of Design"

Accordingly, based on these findings, Markus (1967) suggested adding one more level to the analysis-synthesis-

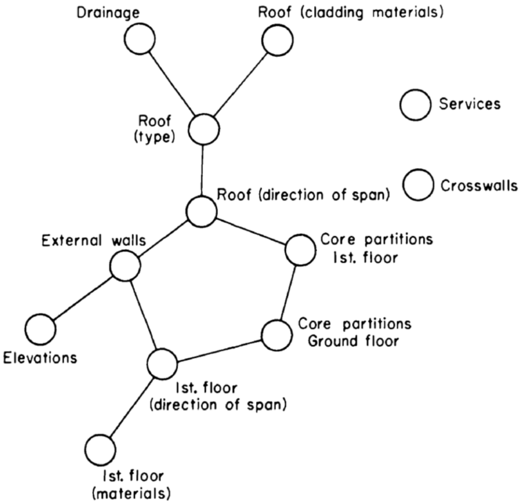


Figure 2.1: Decision graph for the design of a house (source: Luckman 1967)

evaluation process [55]. This higher level consisted of the following three components: preliminary design, sketch design, and detailed design. Each component received input and produced output information to the next component (see Figure 2.2).

[55]: Markus (1967), "The role of building performance measurement and appraisal in design method"

Furthermore, in his seminal book The Sciences of the Artificial, Herbert A. Simon (1969) presented design as a search process of the 'solution space' directed at finding a satisfactory solution [5]. Simon's (1969) approach defined design as a rational problem-solving paradigm.

[5]: Simon (1969), *The Sciences of the Artificial*

While a major goal of the design process is the search for a solution, this process is also based on iterations [56]. Said differently, owing to the existence of complex information dependencies, the iterative design process requires that the design work is repeated over and over again [57]. Through this iterative course, new revisions emerge and are improved or discarded. In line with this notion, the repetitive nature of the design process provides a fundamental structure for the development of design process models [7].

[56]: Kline (1985), "Innovation Is Not a Linear Process"

[57]: Smith et al. (1998), "Experimental Observation of Iteration in Engineering Design"

[7]: Takeda et al. (1990), "Modeling design processes"

From a different perspective, Come, Smithers, and Ross (1994) argued that, rather than a search process, the

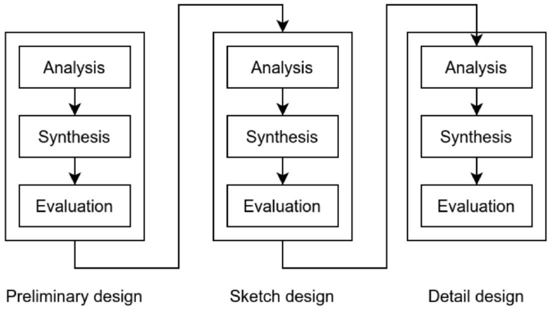


Figure 2.2: A commonly accepted two-dimensional model of the design process (source: Reproduced after Cross 1977)

[58]: Corne et al. (1994), "Solving design problems by computational exploration"

design process is an exploration [58]. The authors proposed that, while a search would generate solutions for well-defined problems, an exploration can derive a problem and propose relevant solutions from an ill-defined problem.

[8]: Maher et al. (1996), "Formalising Design Exploration as Coevolution: A Combined Gene Approach"

Extending this point, Maher, Poon, and Boulanger (1996) presented an evolutionary design process based not only on the exploration of the solution space but also on the evolution of the problem space [8] (see Figure 2.3). The proposed co-evolution model develops both problem requirements and solution candidates until a satisfactory fit can be achieved. In applied studies, the co-evolution model was found to be useful in describing the design process [9, 59, 60].

[9]: Dorst et al. (2001), "Creativity in the design process: Co-evolution of problem-solution"
[59]: Maher et al. (2003), "Co-evolution as a computational and cognitive model of design"
[60]: Wiltschnig et al. (2013), "Collaborative problem-solution co-evolution in creative design"

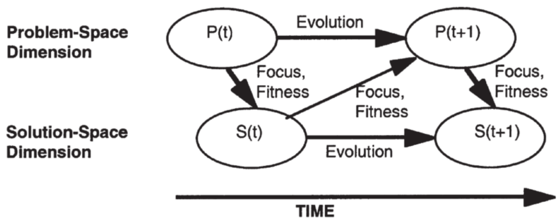


Figure 2.3: Problem-design exploration model (Maher, Poon and Boulanger 1996)

Furthermore, Maher, Poon, and Boulanger (1996) implemented their co-evolution model with a genetic algorithm, which is a specific kind of AI. A genetic algorithm consists of the following two components: a generator that creates solutions and a fitness function that is used to select the best solutions. Jointly, these two components implement the Darwinian logic, suggesting that this model could be meaningfully used with a different method, such as crowdsourcing.

The study of creativity deals with the process of producing creative designs and people's evaluation of those designs to determine their level of creativity. In the literature, creativity has been conceptualized into the following two types: (1) 'individual creativity', usually investigated in psychological research on creative behavior, and (2) 'organizational creativity', usually studied as organizational behavior and collective intelligence [11].

[11]: Maher (2011), "Design Creativity Research: From the Individual to the Crowd"

The individual design process is a stream of expressions, verbal statements, gestures, and actions gathered during or after a design session. In contrast, the collaborative design process is a protocol based on a stream of expressions, actions, and so forth *gathered during a joint design session*. In essence, the collaborative process is based on communication between the participants [11]. The analysis of this collective stream provides a foundation for the study of collaborative creativity by making relations between communication and changes in the designed artifact.

Overall, the creative crowdsourcing process consists of two parts. The first is the individual part, where designers create a design artifact that addresses a specific design problem [11]. The second part is a group discussion held between participants or experts and may include voting to select the most appropriate design. The research of the first part of the process, where the designer works offline and does not share the design protocol, is similar to cognitive research on individual creativity and is difficult to analyze. In contrast, the second part that contains an online discussion or voting is quantifiable, analytic, and researchable [11].

2.2 The Process of Designing Architecture

In order to propose a crowdsourcing method for architectural projects, Royal Institute of British Architects' (RIBA)

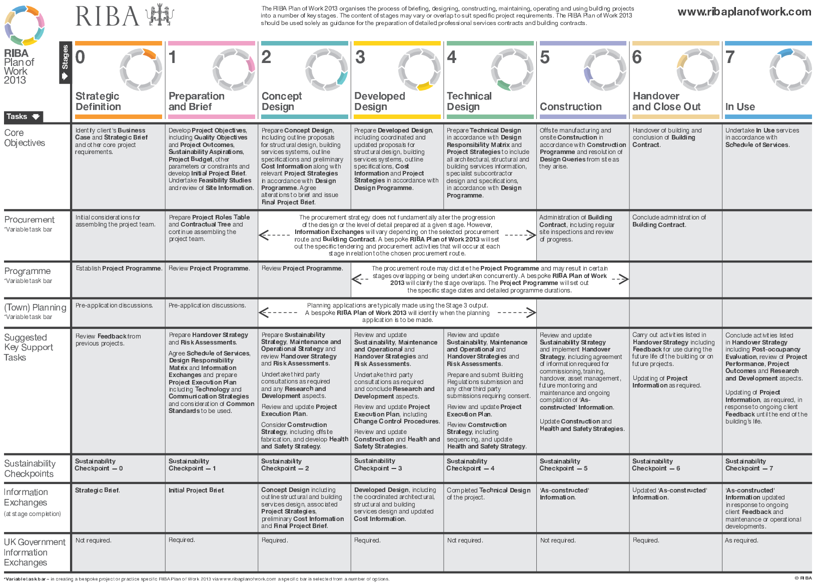


Figure 2.4: RIBA Plan of Work 2013 template

'plan of work 2013' is typically used. RIBA's 'plan of work' is the accepted organizational tool for construction [61]. RIBA's plan outlines the design process with the best practices from the view of the architect (see Figure 2.4). Under the influence of the design method theories [62], this plan was developed to facilitate collaborative teamwork among architects, engineers, and contractors.

[61]: Austin et al. (1996), "A Data Flow Model to Plan and Manage the Building Design Process"

[62]: Broadbent (1988), *Design in Architecture: Architecture and the Human Science*

More specifically, RIBA's 'plan of work' consists of the following eight steps [63]:

[63]: RIBA (2013), *RIBA Plan of Work 2013*

- *Strategic Definition* Identifying the client's business case, strategic brief, and other core project requirements.
- *Preparation and Brief* Developing project objectives, including quality objectives and project outcomes, sustainability aspirations, project budget, other parameters or constraints, as well as the initial project brief and undertaking feasibility studies and review of site information.

- *Concept Design* Preparing a concept design, including outlining proposals for structural design, building services systems, outlining specifications, and preliminary cost information, along with relevant project strategies in accordance with the design program. Agreeing to alterations to the brief and issue the final project brief.
- *Developed Design* Preparing developed design, including coordinated and updated proposals for structural design, building services systems, outline specifications, cost information, and project strategies in accordance with the design program.
- *Technical Design* Preparing a technical design in accordance with the design responsibility matrix and project strategies to include all architectural, structural, and building services information, as well as specialist subcontractor design and specifications in accordance with the design program.
- *Construction* Off-site manufacturing and on-site construction in accordance with the construction program; resolution of design queries from the site as they arise.
- *Handover and Close Out* Handover of the building and conclusion of the building contract.
- *In-Use* Undertaking in-use services in accordance with the schedule of services.

The present dissertation proposes a crowdsourcing model that embeds the creative design process in 'Concept Design', as it is the most influential step in the process of designing an architectural artifact. The 'Strategic Definition' and 'Preparation and Brief' steps are used to define the input of the system. In future research, it may become possible to generalize from our findings into a much broader crowdsourcing workflow, which will cover the remaining steps.

2.2.1 Evaluating Architecture

The question of the value of an architectural artifact has, for centuries, been a source of theoretical thinking;

throughout history, the question of what good architecture is has been answered differently. For instance, Vitruvius (1874) wrote that, for a building to be praised, it must be structurally stable, serve its function, and be aesthetic [1]. Renaissance architect Leon Battista Alberti based his theory on Vitruvius' perceptions and added a geometric mathematical approach that became fundamental to Renaissance art [64]. Over the years, many architects — including the prominent Palladio, Otto Wagner, Luis Sullivan, Adolf Loss, Le Corbusier, Frank Lloyd Wright, Peter Eisenman, Bernard Tschumi, among many others — have offered various aesthetic approaches to the architecture theory. However, despite the variability among these approaches, they all appear to agree that good architecture should be stable, functional, and aesthetic.

[1]: Vitruvius (1874), *The Architecture of Marcus Vitruvius Pollio*

[64]: Alberti et al. (1755), *The architecture of Leon Batista Alberti in Ten Books*

2.2.2 Collaboration in Architecture

In his early writing, Alexander (1977) offered a design theory that rejects the idea of a genius architect who can create a good design without studying the cultural design knowledge system [20]. Alexander's (1977) theory is based on a critical view illustrated by two architects: one who is *unself-conscious* and acting within a rigid cultural boundary, and the modern *self-conscious* architect, who is motivated by ego and ignores previous knowledge and 'way[s] of doing'. Alexander (1977) concluded that the self-conscious architect takes on the almost impossible task of designing a system consisting of thousands of variables, designed over thousands of years by builders who have repaired and improved designs based on mistakes and constraints. Consequently, Alexander's (1977) modern *self-conscious* architect is doomed to failure. The unself-conscious design process can be considered in a systematic manner where the designer is an agent in an existing design-knowledge system.

[20]: Vayssiere et al. (1977), "Notes on the Synthesis of Form"

From this perspective, design is a system based on the pre-existing knowledge created by agents throughout history. Alexander (1977) argued that good architectural work is based on a complex system of designers rather than on the work of a single designer. Overall, there are three types of collaboration in architecture: co-creation, teamwork, and participatory design. In what follows, these three types of collaboration are discussed in further detail.

Alexander's (1977) arguments about the unself-conscious architect highlighted that Architectural design is an integral part of the culture. The techniques and knowledge that have evolved through history are the foundations of architectural education. For example, one method of training a new architect is based on learning and copying from the *Meister* – i.e., the master architects. Therefore, architects inevitably gain new knowledge by embedding the knowledge and creations of others and adapting them into their own creations [65]. Likewise, Dawkins (1976) proposed the 'Memes' model based on a biological metaphor where ideas are copied from one person's brain to that of another person, thus evolving and undergoing a transformation [66]. This kind of collective creativity is referred to as 'Mimetic,' i.e., a creative collaboration in the broadest cultural sense [65].

[65]: Leach (2016), "The Culture of the Copy"

[66]: Dawkins (1976), *The Selfish Gene*

Another type of architectural collaboration is the teamwork of design professionals, particularly in larger projects. Such teamwork emerges in the architects' studio among architects, designers, engineers, planners, and contractors [67]. This kind of collaboration is characterized by the emergence of highly technical and complex business workflows and business models, such as Design-Bid-Build, Design-Build, Construction Management at Risk, Integrated Project Delivery [68], and Multi-Party Agreements [17] that require a specific method for negotiating architectural design across different domains [69].

[67]: Olsen et al. (2014), *Collaborations in architecture and engineering*

[68]: Eastman et al. (2011), *BIM Handbook: A Guide to Building Information Modeling for Owners, Managers, Designers, Engineers and Contractors*

[17]: Smith et al. (2017), *Leading Collaborative Architectural Practice*

[69]: Haymaker et al. (2000), "Filter mediated design: Generating coherence in collaborative design"

Finally, the third type of collaboration refers to the creative connection between the architect and the stakeholders or end-users. There is extensive research on the relationship between the designer and the user, and the different approaches can be mapped on a scale between two major approaches [24]. The first of these major approaches is the 'user-centered' design – a design process where users

[24]: Sanders et al. (2008), "Co-creation and the new landscapes of design"

are the central factor of the design. Requirements are collected using various methods at the initial stage of design. The second major approach is the 'participatory design' approach, where the user is seen as a partner and provided with the knowledge and tools to participate in the design — for example, using Alexander's (1977) pattern language.

In architecture, there is a growing interest in the relationship between architects and stakeholders, particularly in urban planning, where end-users are not entrepreneurs. Specifically, there have been many arguments that the disconnection between architects and end-users in large-scale urban projects is a major factor that leads to inferior urban design [21, 70, 71].

[21]: Jacobs (1961), *Death and life of great American cities*
[70]: Alexander et al. (1977), *A Pattern Language: Towns, Buildings, Construction*
[71]: Salingaros et al. (2011), *P2P Urbanism*

Accordingly, in recent years, the newly developed participatory design methods have become more popular [72–74]); however, these methods still have many limitations [75].

[72]: Habraken et al. (1988), "Concept design games"
[73]: Forester (1999), *The deliberative practitioner: Encouraging participatory planning processes*
[74]: Alexander et al. (1975), *The Oregon Experiment*

[75]: Robertson et al. (2012), "Challenges and Opportunities in Contemporary Participatory Design"

The present dissertation examines the potential of crowdsourcing in architecture to improve teamwork in a way where the participants do not have to be a conventional 'team' but are rather a 'crowd' of professionals and hobbyists from whose combined effort architectural design emerges.

Architectural crowdsourcing also applies to other types of collaboration, such as participatory design. The crowdsourcing process proposed in the present study also involves a process of selecting the most appropriate designs and providing feedback. This process could also be accessible to stakeholders without professional expertise. As opposed to existing participatory planning methods, crowdsourcing can help gather valuable information from stakeholders on a broad and accurate scale without politics, effort, or inefficiency.

2.3 Tacit Knowledge

Knowledge theories provide a foundation for a discussion on expert knowledge and non-expert knowledge, both of which are important to determine the roles of architects and stakeholders in the architectural crowdsourcing process.

In his 1958 book *Personal Knowledge*, Michael Polanyi, a chemist interested in scientists' perception of scientific knowledge, claimed that the absolute objectivity of scientific perception is an illusion. According to Polanyi (1958), the way scientists know something is personal (i.e., embedded in a human being), and this view challenges the approach that scientific knowledge is objective. Being personal is not a flaw of knowledge, but instead is an essential and inseparable part of knowledge.

In his next book, *The Tacit Dimension*, Polanyi (1966) introduced the idea that personal knowledge has a tacit dimension that cannot be easily or systematically expressed. Polanyi (1966) argued that, since "we can know more than we can tell" (p. 4), it is possible to transfer our knowledge partially [77].

[77]: Polanyi (1966), *The Tacit Dimension*

This idea was extended by Nonaka (1994), who proposed to differentiate *explicit knowledge* from *tacit knowledge*. According to Nonaka (1994), explicit knowledge can be expressed in words, numbers, symbols composing drawings, mathematical functions, and codes and be stored in books and computers (see Figure 2.5). With regard to tacit knowledge, it can be transmitted through language, but this transfer is limited because it is sometimes difficult to explain why things are done in a specific way or why they work [78]. Nonaka (1994) also suggested a dynamic model where tacit knowledge can become explicit knowledge and vice versa by socialization (tacit to tacit), externalization (tacit to explicit), combination (explicit to explicit), and internalization (explicit to tacit).

[78]: Nonaka (1994), "A Dynamic Theory of Organizational Knowledge Creation"

For instance, stonemasonry is a profession where the knowledge of sculpting stone is tacit [79]. When asking a stonemason why he decides to hit a stone with the chisel

[79]: Shaked et al. (2020), "Autonomous in craft"

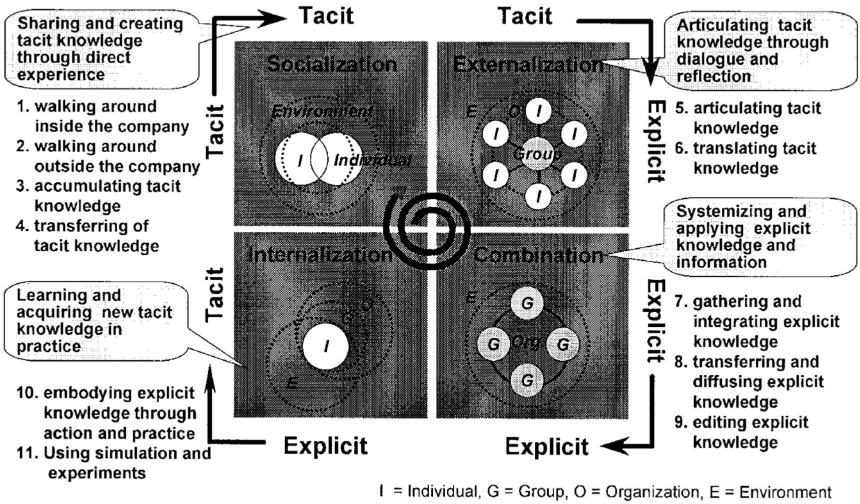


Figure 2.5: SECI model of knowledge creation developed by Nonaka (1994), Nonaka et al. (1994), Nonaka & Takeuchi (1995), Nonaka et al. (2000), Nonaka & Toyama (2003)

in a specific angle and power, he would not be able to explain why he did this action; the only thing that he can say is that he did what was needed. In the same sense, it is difficult, if not impossible, to learn stonemasonry by lingual explanation; instead, it is learned by experience and socialization.

In the same sense, in this dissertation, we argue that the knowledge of the architect required for designing buildings is mostly tacit. This claim is supported by the socialization and experience-based education required for becoming an architect, as suggested in Nonaka's (1994) model. In the past, the art of designing and building structures was learned by close apprenticeship where young aspiring architects would work and learn from master architects [80]. In the 19th century, architecture began to be taught in Ecole des Beaux-Arts through a 'Studio' (atelier), where students would experience creating architecture under the direction of a master-teacher — an experienced architect. Since then, the studio concept spread to other architecture schools as an essential tool for acquiring tacit expert architectural knowledge [80].

[80]: Goldschmidt et al. (2010), "The design studio crit: Teacher-student communication"

Therefore, stakeholders in an architectural project may lack the knowledge of building design unless they are experienced architects or builders [40]. This knowledge gap may limit the possibility of stakeholders to generate or articulate architectural designs. However, stakeholders may hold a deep knowledge of the place and culture, both of which are critical factors in the evaluation of the quality of a design solution. Moreover, stakeholders may identify good architecture because of their life-long experience of living and using buildings. In summary, tacit knowledge provides a theoretical foundation for the participatory aspects of the crowdsourcing model. It determines and clarifies which participants could perform which micro-task based on their apriori knowledge.

[40]: Dortheimer et al. (2020), "Open-source architecture and questions of intellectual property, tacit knowledge, and liability"

2.4 Crowdsourcing Methods

In crowdsourcing research, a distinction is drawn between two types of crowds: laypeople and experts [81]. The former type of crowds frequently refers to large novice (non-expert) crowds to achieve expert-level results by using simple tasks [82]. However, in the present dissertation, we focus on the second type of crowds — namely, a smaller crowd of experts with deep knowledge in their field who collaborate to generate complex products [81]. Since architects hold a unique kind of knowledge and experience required to solve design problems and communicate architectural solutions, we see architects as experts in the present dissertation. However, we also argue that stakeholders' knowledge should also be considered as significant, as non-experts may hold a deep knowledge of the sense of place, environment, and culture, all of which are essential for the evaluation of the quality of design solutions.

[81]: Kittur et al. (2013), "The Future of Crowd Work"

[82]: Retelny et al. (2014), "Expert Crowdsourcing with Flash Teams"

We review several crowdsourcing systems, with a particular focus on research that tries to capture human creativity in creative and complex tasks, such as graphic design, software development, article writing, and so forth. The review focuses on works relevant to our research. As we shall see, each method tries to implement

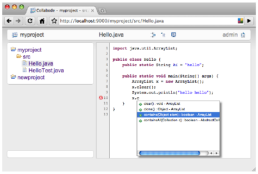


Figure 2.6: The Collabode web-based IDE allows multiple simultaneous editors to work together by Goldman et al. (2018).

a different mode of operation, and each has its own advantages and disadvantages. Our work integrates the essential ideas from these methods into a crowd-based architecture design process.

One of the earliest examples of collaborative programming was 'Collabode', a program implemented through an online programming environment [83]. 'Collabode' is similar to the Wiki software that operates many crowdsourcing websites like Wikipedia, with the exception that all participants in 'Collabode' see the same code changes in real-time. While this system was mostly successful in generating computer code, some participants found the real-time interface to be somewhat confusing since the code changed frequently. We conclude that online collaborative work should be an asynchronous process — i.e., such that it supports the independent and parallel execution of tasks.

[83]: Goldman et al. (2011), "Real-time collaborative coding in a web IDE"

Figure 2.7: The CrowdCode environment and the Write Function microtask by Latoza et al. (2014).

Furthermore, Latoza, Towne, Adriano, and van der Hoek (2014) developed a crowdsourcing method that breaks down the complex and creative task into micro-tasks [84]. Unlike Goldman et al.'s (2011) approach, Latoza et al.'s (2014) system provided tasks in a way that was separated from the general context of the software, and each participant was required to perform a well-defined computer programming task. The researchers outlined a variety of individual structured tasks, such as function specification, automated test writing, and so on. Some participants of Latoza et al.'s (2014) study reported that the nature of the well-structured tasks, without the participants' exposure to the broader context, limited their performance. Likewise, other participants stated that they wanted to communicate with the applicant or to share knowledge, request clarifications, and ask questions regarding code written by others. A similar conclusion was made by Donatcheva, Gerber, and Lewis (2011), who suggested that, in order to increase the quality of products through aggregative creative crowdsourcing, participants should have a communication channel with other participants, rather than only the one with the requester [85]. This evidence suggests that the crowdsourcing project par-

[84]: LaToza et al. (2014), "Microtask programming: building software with a crowd"

[85]: Dontcheva et al. (2011), "Crowdsourcing and creativity"

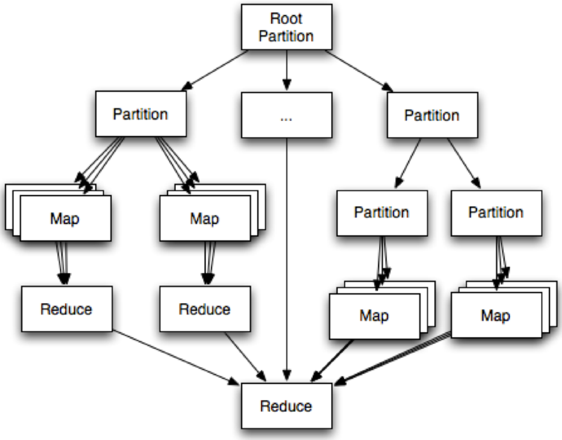


Figure 2.8: CrowdForge's partition, map, reduce workflow by Kittur et al. (2011).

ticipants who receive a well-defined micro-task should, along with communication with the requester, also understand the global scope of the project.

Another general-purpose framework to solve complex problems using crowdsourcing is CrowdForge [86]. The system implements partition, map, and reduce components. First, the participants divide the problem into smaller, manageable sections. Next, the participants map out possible solutions to each part of the problem. Finally, through reduction, the participants unify and choose the best solution. This process makes it possible to write articles, make purchasing decisions, and compose reports on scientific subjects. In the present dissertation, we implement the map-reduce part in our model as the method of exploring the solution space. We also experimented with the partitioning of an architectural design problem.

[86]: Kittur et al. (2011), "CrowdForge Crowdsourcing Complex Work"

2.4.1 Crowdsourcing of Design

Sun, Xiang, Chen, and Yang (2015) developed a method and software, called 'Sketchfun', to solve product design problems by collecting solutions expressed through sketching [87]. The design task was published as text, and

[87]: Sun et al. (2015), "Collaborative sketching in crowdsourcing design: a new method for idea generation"

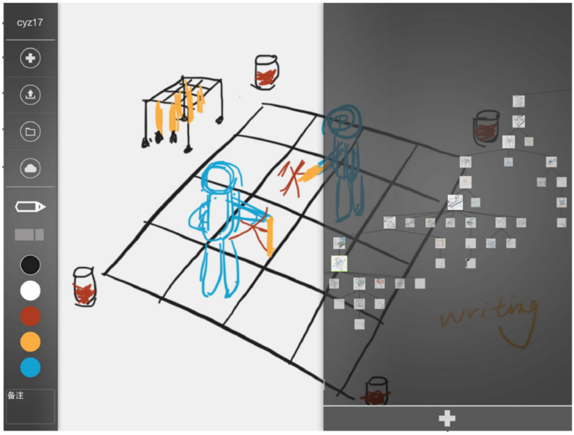


Figure 2.9: Sketchfun's user interface that allows participants to select and improve a sketch from the artifact tree by Sun et al. (2011).

the participants needed to draw the solution. The system then gathered the ideas and rewarded the best ones. Ideas were arranged in 'idea trees', with similar ideas merged into the same branch. The participants could then continue developing the existing branches and created new idea branches. In this way, the system conducted a competition, allowing the participants to see the ideas of others, improving and thus promoting the evolution of their own ideas or those of others. In our model, we adapted the following two features of 'Sketchfun' to our model: (1) using sketches to provide fast design solution representations; and (2) the concept of the idea tree to organize and develop ideas (see Chapter 5 for the use of these features in our experiments).

Furthermore, Yu and Nickerson (2011) experimented with a crowdsourcing evolutionary genetic algorithm based on a biological metaphor where new solutions emerge from merging ideas; random mutations are generated, and a natural selection mechanism is applied to choose the most appropriate solutions [88]. The solutions are perceived as populations that develop over generations, and the optimal solution develops through mutations and combinations. In Yu and Nickerson's (2011) experiment, the participants were required to design an alarm clock. Other participants rated the produced designs and evaluated their levels of functionality and innovation. Another

[88]: Yu et al. (2011), "Cooks or Cobblers? Crowd Creativity through Combination"

Figure 2.10: Examples of the best and worst floor plan designs generated by non-expert workers in Wu et al. (2014).

group of the participants combined the highest-rated designs, creating new ideas that joined the existing idea population, and so forth. Yu and Nickerson (2011) found that the designs of later generations were perceived as more original and functional. From this evidence, we presume that merging different ideas may have the potential to be more creative. Accordingly, in the present dissertation, we conducted some merging experiments where various strategies were used (Sections 5.3.2, 5.4.6, and 5.4.7). From these experiments, we developed an artifact merging task that outputs a 3D model (Section 5.2.5).

In another relevant study, Wu, Korney, and Grat (2014) implemented a crowdsourcing method using non-expert workers to design the placement of furniture pieces in a room. The participants were asked to use online drawing software [89]. In Wu et al. (2014) study, the participants received the following two types of tasks: (1) to create an original design and (2) to evaluate and rank the existing designs. Overall, Wu et al.'s (2014) study was the first to use inexpert participants in architectural design tasks, and the results suggested that stakeholders if provided with simple tools, can actively participate in design. Although the present dissertation study focuses on design tasks performed by professionals, we also explored if non-expert participants can perform design tasks.

[89]: Wu et al. (2014), "Crowdsourcing Measures of Design Quality"

2.4.2 Quality Assurance in Crowdsourcing Systems

The quality of the output generated in a crowdsourcing process is another critical challenge due to the large scale of work and the low transaction cost limiting the effort invested by workers, which results in low-quality output [90]. Substantial previous research has been devoted to identifying workers who provide low-quality work or misuse the system and may generate "garbage data" [91]. A special aspect of the crowdsourcing model developed in the present study is the quality evaluation method for the architectural artifacts. Such a method is necessary not only to identify invalid artifacts but also to select the best artifacts as an integral part of the design process.

[90]: Kittur et al. (2008), "Crowdsourcing user studies with Mechanical Turk"

[91]: Shah et al. (2014), "Double or Nothing: Multiplicative Incentive Mechanisms for Crowdsourcing"

The model presented in this dissertation employs several quality control strategies based on the concepts developed in previous research. Three of these strategies are discussed below in further detail.

First, *fault-tolerant task* is a strategy used to make a micro-task less sensitive to low-quality work. For example, Kittur, Smus, and Kraut (2011) applied the partition-map-reduce algorithm to generate redundant tasks; then, the quality of the results was rated, and the highest-rated result was selected. In the present study, we apply this strategy by systematically assigning multiple identical tasks to different participants. Later, the outcomes are processed and analyzed in different ways to generate and select high-quality output.

The second strategy is *instructions optimization*, which means that the task instructions are tested and improved to achieve the participants' optimal performance. In one relevant study, Chandler and Kapeler (2013) optimized the performance of crowd workers by changing a task description and measuring the worker's performance [92]. The results of Chandler and Kapeler's (2013) experiment showed that highlighting the significance of a task resulted in a higher participation rate, better quality, and larger volumes of work. In our experiments, multiple task formats, descriptions, and graphic user interfaces were

[92]: Chandler et al. (2013), "Breaking monotony with meaning: Motivation in crowdsourcing markets"

tested while measuring the resulting task completion rate and output quality. Based on the results, the task design was improved, leading to better output results.

The third strategy which we employ in this dissertation is *manipulating incentives*, which implies that the quality level depends on the compensation model. For example, Shaw, Horton, and Chen (2011) found that explaining the meaning and importance of the task to crowd workers increased the performance by raising the participation and higher quality results [93]. Likewise, Latoza et al. (2014) found that giving points as ratings to participants in programming tasks had a large positive effect on the participants' motivation and work, especially given the relatively high payment they received. In the present dissertation, we adopted these approaches in our model and provided the "project brief" as an input of each micro-task to ensure that the study participants understood the meaning and importance of the task. We also awarded workers with points on the completion of tasks.

[93]: Shaw et al. (2011), "Designing Incentives for Inexpert Human Raters"

2.4.3 Design Review Crowdsourcing Methods

In the architectural crowdsourcing process proposed in this study, the participants were asked to express their opinions on the various artifacts by answering multiple questions. This generated feedback helped the designers to improve the designs. Previous studies suggested that the feedback given to participants improves the quality of the generated solutions [94].

[94]: Wooten et al. (2017), "Idea Generation and the Role of Feedback: Evidence from Field Experiments with Innovation Tournaments"

'Open feedback' is the most common and simple method of generating design critique using crowdsourcing [95]. For example, websites that focus on creative work provide an open feedback form for users to provide *unstructured* feedback[1]. The unstructured nature of this approach generates low-quality feedback as compared to the one afforded by structured methods [95].

[95]: Greenberg et al. (2015), "Critiki: A scaffolded approach to gathering design feedback from paid crowdworkers"

1: For example, behance.com, and Dribble.com

For instance, CrowdCrit applied a 'structured' approach that allowed both non-expert and professional crowd workers to generate feedback on graphic designs [96]. The

[96]: Luther et al. (2015), "Structuring, Aggregating, and Evaluating Crowdsourced Design Critique"

system collected feedback from users through structured forms, which not only contributed to organizing the information but also enabled free text input that gave users the freedom to express complex ideas. The system was found to be effective in obtaining valuable feedback that was approaching expert-level feedback.

Another example of a crowdsourcing design feedback system is 'Critiki' [95]. The system is based on the 'scaffolding' approach — a method where the participants are guided through smaller subtasks in a sequence that supports them in completing a larger task. This approach differs from the previously mentioned 'divide-solve' approaches, where tasks are divided and merged through an algorithm. Greenberg et al. (2015) found out that applying the scaffolding approach generated near-expert results and that the study participants performed much better than when the 'open responses' approach was used.

Based on the evidence reviewed above, in the present dissertation, we assume that feedback can be generated using simple structured micro-tasks. We also expect that the quality of non-expert workers' aggregated feedback could approach that of professional architects. Finally, since inexpert stakeholders in our case may hold important local knowledge, we suggest that their feedback might be even more valuable to the design than that of professional architects.

3 Pilot Experiment

The pilot experiment we started with had the following four objectives. First, we aimed to learn about the crowdsourcing model implemented by 'Arcbazar'[1], a commercial and relatively successful architectural competition "crowdsourcing" website. The second objective was to evaluate a public anonymous architectural competition as a qualitative method for the present dissertation. The third objective was to generate a 'project brief' document that would summarize all requirements and information necessary to produce a design. The fourth objective was to evaluate the performance of freelance architects (recruited through the Upwork platform[2]).

1: see arcbazar.com

2: Upwork.com is a website that helps to find remote freelance workers

The chapter is organized as follows. First, the experiment's method is described (Section 3.1). Then, the generated designs are presented (Section 3.2), and the experiment results are analyzed (Section 3.3). The chapter concludes with some issues arising from the crowdsourcing website and the architectural competition (Section 3.4).

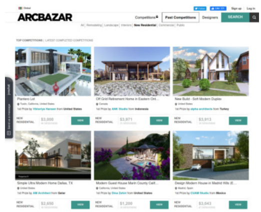


Figure 3.1: Screenshot of Arcbazar's homepage

3.1 Method

The Israel Architects and Urban Planners Association (IAUPA) published an architecture competition for the Safra Square in Jerusalem. This square is known as a problematic public space that is currently not used by the residents [97]. The Jerusalem municipality decided that the square should be improved with a new design and new buildings. The Safra Square compound consists of numerous municipal office buildings; most of these buildings are historical and are under strict preservation. A significant public space is in front of the main office building, which is built on top of a large underground parking structure. At the northern edge, a performance stage was built. Another public space is the Daniel Garden, which is parallel to Yafo street that has beautiful trees and fauna. Between Safra Square, Daniel Garden, and Yafo street, there is a smaller public space with old and dead palm trees that should be removed.

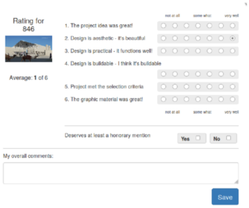


Figure 3.2: Screenshot of Arcbazar's rating user interface

We registered for the competition and, on June 29, 2018, published a project on Arcbazar. The project was titled 'Design a major public square in Jerusalem.' The submission deadline was July 31, 2018. We then submitted the selected Arcbazar design to the IAUPA competition. We functioned as a mediator between the Arcbazar online competition and the IAUPA competition.

Arcbazar provides a platform for online architectural design projects that are mostly based on a single competition cycle. Artifacts are submitted by designers and are later rated by other designers. The 1^{st}, 2^{nd} and 3^{rd} prize winners are awarded a monetary prize, determined by the client. However, the client has full control over the final decision and rating. Later in the experiment, we sent the entries to four Israeli architects who served as a jury and rated those entries independently. The designer, whose design was chosen, compiled it into two panels that were submitted to the IAUPA competition.

Our brief on Arcbazar included the following information: Competition terms (payment, winning, and credit), Competition goals, Project objectives, Possible intervenience points, Tips from the client (a list of suggestions provided in the original competition brief), a list of notable buildings in Jerusalem (including the Israeli Museum, the Supreme court, and others), Current situation, Historic situation, Physical description, 28 selected images of the compound, ten different maps of the area from the municipal GIS website, a 3D area-model and, finally, the required architectural artifacts.

A 3D area model was not made available by the municipality, although multiple competitors requested it. However, we believed that such a model could save our designers much time and raise the participation rate. Through the Upwork platform, an architect from Macedonia was hired to create the model (see Figure 3.3). The model's quality exceeded our expectations and costed roughly 180 USD. The project budget we provided to Arcbazar was 1150

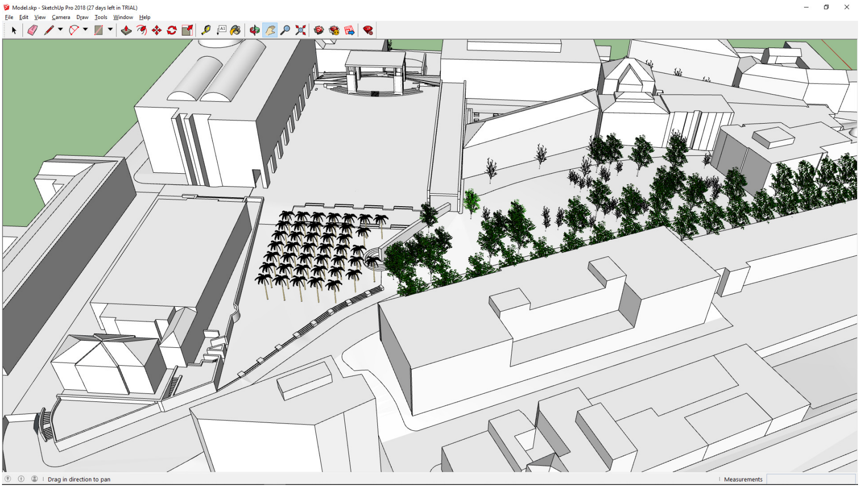


Figure 3.3: Screenshot of the SketchUp Safra Square model that was produced by the Upwork architect

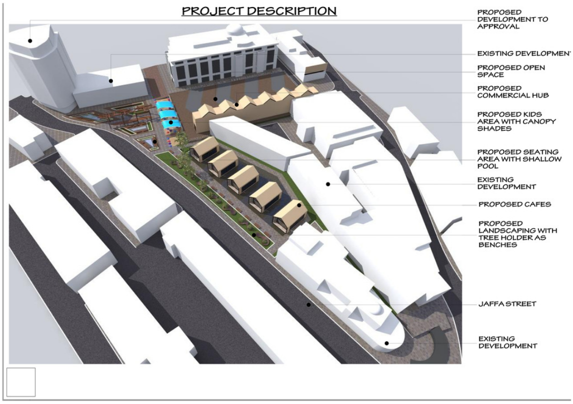
USD. The website distributed it as follows: 600 USD to the 1st place, 300 USD to the 2nd place, 100 USD to the 3rd place, and 150 USD as the website fee.

3.2 Generated Designs

During the time when the project was active on Arcbazar, 25 designers signed up, 13 more saved the project, but only four proposals were submitted. Among these 25 participants, 69.2% were male, 15.4% female, and 15.4% did not disclose their gender. Also, 53.6% of the participants declared that they were professional architects or designers, and 7.7% were students. Finally, 43.6% stated that they received a Master's degree in Architecture, and 30.8% have a Bachelor's degree.

All the entries used the 3D area model we provided. In what follows, further detail on the four submitted designs is provided.

3.2.1 Entry 714



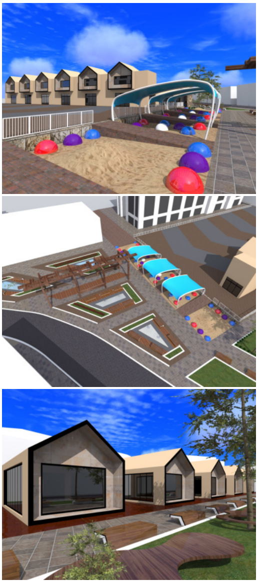


Figure 3.5: Entry 714

This design added a unique 5-gable structure on the eastern edge of the square, suggesting a children's playground instead of the palm square with an elegant pool. It also proposed building five gabled structures on the edge of the Daniel Garden that would serve as a coffee shop in addition to a narrow garden strip.

The proposal's design contradicted the architectural language of the square's 20th century modernist language due to the particular use of gables. It also blocked the traffic route passing through the buildings from the east.

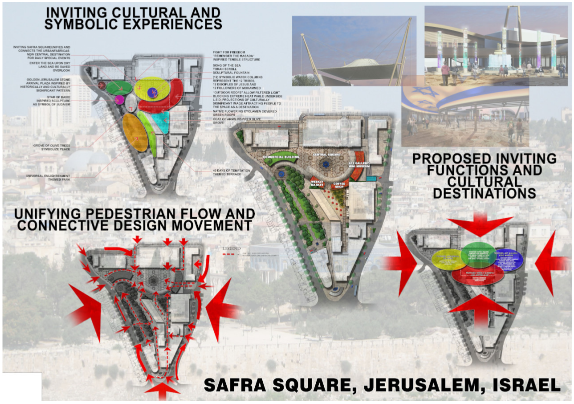
3.2.2 Entry 569

The approach of this design was to add a dramatic structure to the existing performance stage in the northern part of Safra Square. The new structure symbolized the connection between East and West in Jerusalem. Besides, the proposal replaced the Daniel Garden with a landscaped garden that included elements of water. Instead of the palm trees, the design created shaded areas for a coffee shop.

Although the proposal was well developed in terms of design, it did not connect the main square to the street.

Figure 3.7: Entry 569

3.2.3 Entry 866



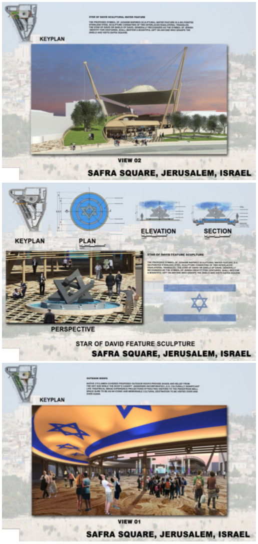
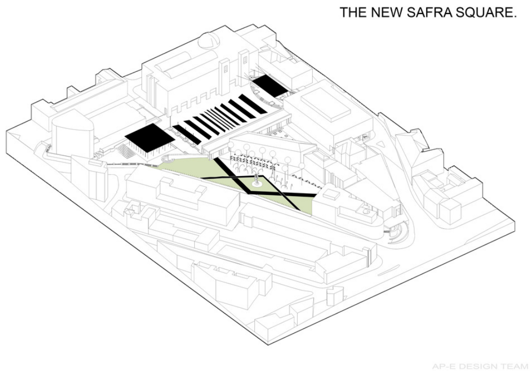


Figure 3.9: Entry 866

The third design suggested a vast tensile structure stretching above the public space, including iconography referring to the Zionist narrative "Masada shall not fall again." It also offered a fountain in the shape of the Dead Sea Scrolls, another fountain in the shape of a broken star of David, and 12 pillars of water representing the 12 Jewish tribes, 12 apostles of Jesus, and the 12 successors of Muhammad. The proposal included the erection of commercial buildings around Safra Square, the preservation of the Daniel Garden.

With all its creativity, this design had numerous concerns. The main problem was the religious and national symbolic iconography, which is problematic in a secular government structure in a politically charged city. Additionally, there was a proportion problem with the size of the tensile structure that rose high above the municipal building roof, and the large ramps blocked the square entrance and the surrounding buildings.

3.2.4 Entry 846



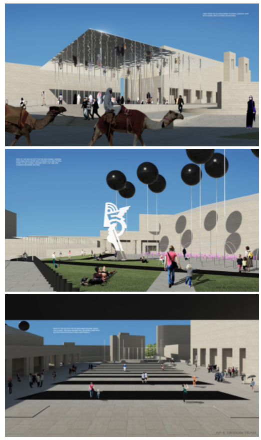


Figure 3.11: Entry 846

The fourth design suggested converting the ground floor of the municipality office building into a commercial space, thus converting a historic building at the rear of the square into a visitor center and turning a historic building that borders with the Daniel Garden wall into a coffee shop.

The design introduced a redesign of the Daniel Garden, turning the Palms space into a roofed market and partially shading the Safra Square. The proposal was relatively limited since it included minimal construction intervention, as the competition sought to expand existing buildings. It also included sizeable black shading and flooring, which would generate tremendous volumes of heat. Uprooting the old trees in the Daniel Garden and using black elements were mistakes.

3.3 Analysis of Designs and Data

After the submission deadline, there was a week-long vote session with the participation of several other crowd workers. The designs were rated with several statements that had to be rated on a scale from 1 to 7. The statements were as follows: 'The project idea was great!', 'Design is

Table 3.1: Distribution crowd votes on Arcbazar and expert rating

Entry	Concept	Aesthetics	Function	Buildability	Graphics	Average	Count	Expert votes
569	7.7	7.6	7.8	8.1	8.2	7.9	13	6
866	6.0	6.5	7.5	7.5	7.4	7.0	8	4
846	6.8	6.3	7.2	7.8	6.3	6.9	5	8
714	5.4	4.0	5.9	7.6	5.9	5.8	5	6

aesthetic - it's beautiful,' 'Design is practical - it functions well!', 'Design is buildable - I think it's buildable', 'Project met the selection criteria' and, finally, 'The graphic material was great!'. After the voting session, entry 569 received the highest average rating from 13 votes, while other designs received lower ratings from fewer voters.

Next, we also asked four expert Israeli architects (e.g., Expert panel) to evaluate the entries on Arcbazar to compare the crowd voting with expert opinion. Entry 846 received the highest score, while on Arcbazar, it was rated only third. Entry 866 received the lowest rating because the experts thought it was a misfit (due to local political reasons), while the crowd workers on Arcbazar rated it in the second place. Table 3.1 shows a summary of the votes on Arcbazar and provides a summary of the expert panel rating.

The divergences between the crowd and the expert evaluation were a problem since the crowd selected a design that was not aligned with our expert panel opinion. We normalized and added the crowd and expert votes and chose entry 846 to be submitted to the competition since it scored the highest.

After the IAUPA competition concluded, we received the competition protocol that included information about the competition, the criteria used by the competition jury to evaluate the designs, and the designs that were selected to advance to the second competition stage. A total of 49 proposals were submitted. The entries were reviewed by the competition's jury using the following criteria: 'Context and continuity,' 'Urban planning,' 'Connectivity', 'Unique identity,' 'Diversity,' 'Location of mass,' 'Site match,' 'Architectural qualities and landscape design,' 'Intensity of the project,' 'Simplicity and modesty,' 'Stability and survivability,' 'Suggested activities in buildings and open spaces,' 'Flexibility,' and 'Applicability'.

Out of the 49 proposals that were submitted to pass the first stage, the competition's jury selected 16 proposals. Our proposal was not selected, and we received no further feedback regarding its quality.

3.4 Conclusions

Arcbazar's project provided four unique designs that each was presenting a different approach. Although the expert panel selected the best design, none of the proposals was regarded as satisfactory. This might have been the result of several factors. First, the project budget was divided by the website between the top 3 designs, which lowered the compensation for the 1st prize by 40%. Accordingly, the remaining designers were not compensated despite their investments in terms of time and effort. Another concern was that the designs had shortcomings related to the limited knowledge of the location, climate, and political situation. Finally, and importantly, we had no influence over the design workflow and were surprised to see the design outcomes. We believe that our intervention would have improved the designs.

On the other hand, through Upwork, we were able to hire a freelance architect who performed well. It was easy to communicate with the freelance architect, and we received the artifacts on schedule; however, a straightforward comparison of the output of Arcbazar to that of a freelance architect is not possible, as the requirements were different.

Unfortunately, the IAUPA competition jury did not provide us with more detailed feedback regarding our design's performance as compared to other submissions, so we do not have comparative data for further analysis. On the other hand, we received performance data from the assembled expert panel and, based on this experience, concluded that it is effective to use an expert panel to evaluate designs in the experimental framework.

Furthermore, we noted more shortcomings of the IAUPA competition. For one thing, participation in the competition is rather costly. Peggy Deamer (2015) estimated that the median work hours for a competition cost 4,000 USD [49]. We validated Deamer's (2015) estimation with architect Ishai Well, who won second place in the IAUPA competition. He acknowledged working on the competition took a month, and we estimated the cost of work of an experienced architect as 10,000 Euro[3]. We also noted that the IAUPA competition winner, Chyutin architects, presented three renderings done by Bloomimages, a global hi-end architectural rendering service company, which we estimated to cost about 10,000 Euro[4]. From this limited study, we can conclude that architects invest much effort and time in competitions: at least a month of work, which equals 10,000 Euro.

[49]: Deamer (2015), "The Guggenheim Helsinki Competition: What Is the Value Proposition?"

3: Work cost includes the worker salary, taxes, benefits, facilities, and other expenses related to employmen

4: Estimation by Ronen Beckerman, an architectural rendering expert

From the competition data, we also learned that the chance of winning was as low as 2.04% since there were 49 competition entries. As mentioned previously, we estimated the participation cost to be 10,000 Euro, while the 1st prize was only 15,000 Euro. Alongside the fact that the prize money was not worth the investment, the competition organizers also gained free innovation-work, including the copyright, worth about 500,000 Euro, without any obligation to realize any design or to order design services from the winner. In the case the organizers would decide to ask the winners to continue the design, the prize money would be deducted from the fee.

This shortcoming also was identified on Arcbazar in our experiment and several previous studies (e.g., [50]). According to Deamer (2015), in high-profile architecture competition, there is a cultural capital gain for the designers; however, Arcbazar's projects generate no cultural effect and, therefore, no cultural capital is gained. Also, the prize money was distributed between the three first places, meaning that the remaining designers were not compensated, and since there are, on average, 10 participants, 70% will receive no payment, while the winner will only receive 60% of the prize. In this respect, Keslacy (2018) argued that Arcbazar is using the competition model to sell professional and creative labor, highly regulated by established themes, for a bargain price without paying the designers.

In conclusion, Arcbazar generated problematic designs and the experts' panel did not reach a consensus on crowd-voting results. We concluded the following:

- The brief was unsuccessful in communicating the complexity of the requirements, local culture, and politics.
- The intervention of stakeholders who have local knowledge is critical for selecting the best design. We suggest that the discrepancy between the crowd-voting result and expert panel may result from the limited crowd's knowledge of Jerusalem's culture and local politics.
- We had no means to intervene in the design process beyond the preliminary brief and answering the participants' questions. Therefore we conclude that a design process that has stakeholder intervention may lead to a better design.
- Arcbazar and the IAUPA competition succeeded in generating different designs; both of them have a flawed and exploitive financial compensation model. This model may exclude professional designers from participating in crowdsourcing design processes and limit the resulting quality.
- The expert panel was a simple and effective way to measure design quality in experiments.

Yet, despite the flaws of the competition crowdsourcing model, we argue that this is an outcome of the specific crowdsourcing model employed by Arcbazar and not a shortcoming of crowdsourcing as a design paradigm. The goal of this dissertation is to identify a desirable crowdsourcing model for architectural design.

In the next chapter, we present a preliminary architectural crowdsourcing model based on the conclusions of our pilot experiment. The four key characteristics of the proposed model are as follows. First, the model explores the solution-space using a micro-design competition, which was confirmed to be successful in generating designs. Second, by suggesting design micro-tasks that require a limited time investment of the designers, it will be possible to compensate them for their effort while also controlling the funds. Third, we suggest micro-tasks that allow non-professional stakeholders with local knowledge to be a part of the design process and influence the outcomes. Fourth, we use an expert panel to evaluate the outcomes of the crowdsourcing process.

Methodology 4

This chapter describes the research methods used in the present dissertation. In Section 4.1, a preliminary crowdsourcing model is specified based on the literature review and the conclusions from the pilot experiment. In Section 4.2, we describe the methods used in the experiments. Section 4.3 describes the tools and software used in the experiments. Based on the experimental results of testing and evaluating the preliminary model, we propose a refined crowdsourcing model in Chapter 6.

4.1 Preliminary Model

This section describes a preliminary crowdsourcing model. The preliminary model is based on our literature review (Chapter 2) and the results of the pilot experiment (Chapter 3). This preliminary model serves as a starting point for the experimental development of the model.

In previous research on creative crowdsourcing models, we identified the following two major micro-tasks:

1. A design workflow, which explores and searches the solution space through competition [87], combination [88], or partitioning of the problem [84, 86].
2. A selection workflow, which limits or merges a number of solutions using 'idea trees' [87] or by voting [86, 88, 89, 98].

[87]: Sun et al. (2015), "Collaborative sketching in crowdsourcing design: a new method for idea generation"

[88]: Yu et al. (2011), "Cooks or Cobblers? Crowd Creativity through Combination"

[84]: LaToza et al. (2014), "Microtask programming: building software with a crowd"
[86]: Kittur et al. (2011), "CrowdForge Crowdsourcing Complex Work"

[86]: Kittur et al. (2011), "CrowdForge Crowdsourcing Complex Work"
[88]: Yu et al. (2011), "Cooks or Cobblers? Crowd Creativity through Combination"
[89]: Wu et al. (2014), "Crowdsourcing Measures of Design Quality"
[98]: Wu et al. (2015), "An evaluation methodology for crowdsourced design"

Our preliminary crowdsourcing model consists of the following three modules.

The first module is the *Design module*. This module consists of multiple *Design micro-tasks*. The module outputs multiple design artifacts.

The second module is the *Selection module* made out of multiple *Selection micro-tasks*. The 'solution space' is limited using rankings provided by the participants. The module outputs the most suitable artifacts.

The third and last module is the *Review module*, which consists of multiple *Review micro-tasks*. In this module, the participants are asked to express a verbal opinion regarding the produced solutions. This feedback serves as a constructive critique of the next design iteration, akin to the "CrowdCrit" system [96].

[96]: Luther et al. (2015), "Structuring, Aggregating, and Evaluating Crowdsourced Design Critique"

The next section provides further detail on the workflow and different micro-tasks, including the required and generated outputs.

4.1.1 Workflow

Creating architectural design requires an evolutionary design process [99]. In the studio, ideas are developed, optimized, and evaluated as regular praxis. However, designing is an exploratory process that is difficult to divide into sub-tasks [100].

[99]: Howard et al. (2008), "Describing the creative design process by the integration of engineering design and cognitive psychology literature"

[100]: Schmitz et al. (2018), "Online Sequencing of Non-Decomposable Macrotasks in Expert Crowdsourcing"

Several creative crowdsourcing methods suggested using an evolutionary logic through 'idea trees' [87] or combination [88]. In this dissertation, we propose an iterative *micro-competition* workflow based on small design tasks. The workflow diagram is shown in Figure 4.1 and is consists of input documents (requirements and area model), micro-tasks (design, selection, and review), stop criteria, and output-document (see below for further detail).

[87]: Sun et al. (2015), "Collaborative sketching in crowdsourcing design: a new method for idea generation"

[88]: Yu et al. (2011), "Cooks or Cobblers? Crowd Creativity through Combination"

Stop Criteria

As mentioned above, while designing is an evolutionary process, it has to conclude with a specific artifact. In the traditional design process, the designer decides that the design is finished when the outcome is a fit solution [5]. In our preliminary model, we agreed that the 'process manager' would decide when the result was adequate, and then the process could be stopped.

[5]: Simon (1969), *The Sciences of the Artificial*

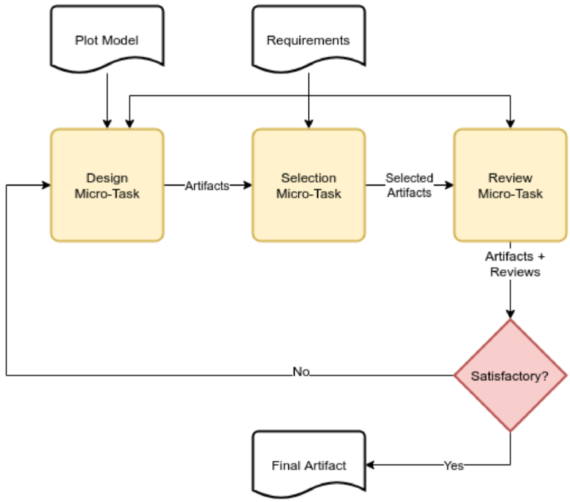


Figure 4.1: Flow diagram of the proposed workflow

4.1.2 Protocol Objects

The information transferred between modules is defined by the following two types of objects: (1) the brief object and (2) artifact objects.

Brief Object

A brief is an object that includes the products of the 'Strategic Definition' and 'Preparation and Brief' steps in the RIBA's 'plan of work.' This includes project objectives, quality objectives, project outcomes, sustainability aspirations, budget, constraints, project brief, and 'site information.' The brief also has to include a file with a 3D area model of the surrounding buildings, thereby helping the participants to create designs that better fit the environment.

Artifact Object

Artifacts are objects that include an architectural design or a digital representation of a sketch. For instance, an artifact may have a SketchUp model file and include some images of renderings of the model. An artifact also includes all generated review objects concerning the specific artifact.

4.1.3 Modules and Micro-Tasks

As mentioned, the iterative workflow consists of the following three types of modules: (1) the design module, (2) the selection module, and (3) the review module.

Each task description is made out of (1) task process, (2) input that is provided, (3) desired output, (4) kind of participants that should perform the task, and (5) an output validation method.

Design Module and Micro-Task

The first step was to find design alternatives, a process known in the literature as *solution space exploration*. The participants were asked to produce artifacts using the 'SketchUp' software; therefore, this task requires professional skills.

SketchUp is a popular and straightforward, and multi-purpose CAD software. The user interface of SketchUp is simple but has many sophisticated geometric features that allow users to start quickly while supporting professional features. SketchUp also has a free version provided as a browser application connected to a free file collaboration service. The simplicity and high availability of SketchUp make it a useful tool for our experiments.

Process The participants were asked to read the project requirements. Then, they were required to generate an architectural artifact based on the requirements. Upon completion of the aforementioned two steps, they uploaded a digital file to the system, and the task was completed.

Input The participants received the brief. In later iterations, the input included artifacts originating from previous iterations. These artifacts also included the critique generated during the review micro-task.

Output The participants produced and uploaded a CAD model to the application for storage and subsequent use during the next step.

Participants This task required professional architects with knowledge of SketchUp. We hypothesized that architectural design and CAD knowledge are the architects' *tacit knowledge* acquired through previous experience.

Validation This task was not structured and required validation. The CAD model files should be valid SketchUp files containing an architectural artifact. First, only SketchUp files (identified by SKP file extension) were accepted. Second, the selection micro-task displayed the artifact to a human who would dismiss and reject artifacts that were not architecture or did not fulfill the requirements.

Selection Module and Micro-Task

After an array of artifacts was generated in this task, the most fitting artifacts were selected. The participants were asked to rate the artifacts based on the requirements and personal evaluation. As mentioned previously in Subsection 2.2.1, there are many approaches to evaluating architecture. For the preliminary model, we preferred a simple evaluation method. We used a scale (between -1 and 1) and requested a rating for the three Vitruvian conditions of good architecture: function, stability, and aesthetics.

Process The participants were asked to read the project requirements. Next, they were presented with a CAD model, video, and various images of the artifact. Then, the participants were asked to answer various questions, such as "Is this model aesthetic?", "Does this model meet its requirements and is functional?" and" Do you think that this model is structurally stable?". The participants selected a numerical score from a range with textual labels. For example, the question "How do you evaluate the design quality?" would have the following answers 'Low' (-1), 'Average' (0), and 'High' (+1). After a sufficient number of answers were collected, the answers were analyzed, a score was calculated for each artifact, and the lowest-rated artifacts were discarded.

Input The input for this task was the brief and an array of artifacts.

Output After all participants finished rating all artifacts, and a score for each artifact was calculated, the output was provided as a trimmed array of top selected artifacts.

Participants Contrary to the design task, this module did not require professional knowledge; instead, it presupposed the understanding of the location and project.

Validation This was a well-structured task where participants could choose only from a limited list of options. The default answer was 'none'; therefore, the participants were not forced to take a stand if they did not have one.

Review Module and Micro-Task

This module generated constructive feedback that provided the participants with insights and directions subsequent design steps. In previous research, feedback-system generated specific observations for graphic designs using visual markers on the designed object, categorized questions, and free text proved to be effective [96]. Our crowdsourcing model adopts a similar functionality in this module—the one in which the participants viewed the artifact and had a web form to provide a critique. In that way, the participants were able to review and communicate their critique to other participants.

Process After the participants read the requirements, a list of artifacts was presented on the screen. The participants were asked to answer the following questions: "What do you like in this design?" "What would you like to change in this design?" and "What would you remove from this design?"

Input The input for this task was the brief and one artifact.

Output After the participants provided feedback, it was 'attached' to the specific artifact.

Participants The participants were either professional designers or stakeholders and other non-expert workers.

Validation This task was well-structured and required the participants to write a minimum-length answer to each question. The process would not continue unless each artifact had at least four reviews.

4.2 Experiment Design

This section describes the research method applied to answer the research questions. In the central part of the study, two series of experiments were performed as workshop courses.

4.2.1 Research Through Design

Human-computer interaction (HCI)-based research investigates the design and use of computer technologies, with a particular focus on interfaces between people and computers. While the present thesis focuses on architectural design, we apply several pertinent HCI research methods — namely, research through design (RtD) [101]. Contrary to the applied sciences, design seeks to solve an ill-defined problem [5], which makes design a unique discipline [102, 103].

[101]: Frayling (1993), "Research in Art and Design"

[5]: Simon (1969), *The Sciences of the Artificial*

[102]: Cross (1982), "Designerly Ways of Knowing : Design Discipline"
[103]: Cross (2001), "Designerly Ways of Knowing: Design Discipline Versus Design Science"

Since RtD is an inquiry process focused on a product's manufacturing, service, environment, or system, the knowledge gained in this process can be implicit and reside almost entirely within the resulting artifact. Additionally, RtD is not a formalized approach.

While there have been active debates around appropriate design research methods [104], in the present study, we relied on Zimmerman, Forlizzi, and Evenson's (2007) criteria of high-quality RtD *process*, *relevance*, *invention*, and *extensibility*.

[104]: Zimmerman et al. (2010), "An analysis and critique of Research through Design"

The critical element of interaction design research is documenting the *process*. In principle, two designers are not expected to produce the same design to solve the same problem. Furthermore, to reproduce the process, designers should be provided with details about the process, methods, and rationale. Design research should also fulfill the criterion of *relevance* that frames the work within the real world and foregrounds the designer's preferences of the design state. In addition, the research contribution should be novel and constitute a significant *invention*. Finally, the knowledge should be *extensible* to allow building upon it. Therefore, it is crucial to describe the design and research process, experiments, pertinent questions, and results that underlie the conclusions.

4.2.2 Experimental Workshop

The first set of experiments (Experiments 1 - 15) was designed to explore and test the initial feasibility of the preliminary model using the experimental modular software. The second set of experiments (Experiments 16 - 26) focused on the analysis and conclusions from the first set. The enhanced model was tested using a new software program developed specifically for that purpose.

Two one-semester courses were designed as a practical workshop for second- to fifth-year architecture students at Tel Aviv University. These students performed as crowdsourcing workers. To minimize the potential effect of the workshop's academic requirements on the results, the students were graded based on their attendance and the number of completed micro-tasks.

Each class included several experiments. The completion time of each task varied, so the experiments had different duration. For instance, a selection micro-task experiment lasted 10 minutes, followed by a review micro-task experiment (15 minutes) and a design micro-task experiment (60 minutes). On completing the tasks, the study participants were asked to complete a questionnaire about their experience of performing the experiments (5 minutes).

The first workshop in the semester started in October 2018, and each class lasted 90 minutes. The workshop included 16 students, three students from the fifth year, two from the fourth year, one from the third year, and ten students from the second year. Therefore, most of the students had minimal professional experience.

The second set of experiments took place in the next semester in February 2019, and a total of seven students were enrolled. We continued experimenting while comparing different experiments as the control group. Also, the artifacts were presented to experts who evaluated the works and measured their quality.

In addition, we recruited professional architects from the 'Upwork' platform to participate in the experiments. In doing so, our initial motivation was to promote the design process with a focus on new experiments. However, the inclusion of freelancer architects also provided us with valuable insights since we were able to validate the method with crowd workers and professional architects. Their feedback helped us gain confidence in workflows that the students failed due to their lack of experience.

4.2.3 Experimental protocol

The present study aims to test, evaluate and compare experimental crowdsourcing systems. In experimental design, we followed a principled experimental protocol based on HCI experimental practices [106].

[106]: Gergle et al. (2014), "Experimental Research in HCI"

Before each experiment, we established the aim, hypothesis, goals of each experiment—for instance, to test if a specific micro-task provides the expected output and how this output compares to the output of previous experiments.

During the experiment, the participants used the software and performed the micro-tasks randomly or based on system logic. Upon completion of the experiment, the participants responded to a short survey about their experience. The survey results included qualitative and quantitative data.

After each workshop class, all data were recorded. The course of the experiment and results were analyzed and documented. The generated artifacts were evaluated by four expert architects who rated their general *quality score*. Since this evaluation was subjective, we averaged, normalized, and rounded expert scores to mitigate deviations. Accordingly, the obtained quality scores were deemed to be good indications of both the design micro-task quality and the selection micro-task performance.

Moreover, we reviewed each produced artifact and analyzed it qualitatively by describing it in architectural terms. We also analyzed the generated reviews, identifying and all issues that emerged in the process.

In the next step, the results of each experiment were compared with those obtained in previous experiments; in some cases, comparisons were additionally made with the performance of an internal control group that performed another micro-task. The quantitative results were analyzed using different factors, including participant experience, task execution time, text length, repetition rate, expert quality score, and participant quality evaluation. We also performed correlation tests to identify the factors that could have affected the results.

Finally, we examined the results of the experiments, including the input of each micro-task, corresponding quality and accuracy of the results, ease, and conformity of the products with the requirements. The aims and course of the next experiment were revised and updated based on the conclusions and new questions that arose from the experimental results.

4.2.4 Validity of the Methods

Since most of our experiments were performed by the same group of students, methodological risks related to the pedagogical context have to be addressed.

First, there could have been a risk that the students would not select the best artifacts to get better marks at the workshop. To mitigate this risk, we told the participants that their scores depended exclusively on their active participation in the class. That is, the students were told that there would be no penalty for those who attended the class but failed to perform micro-tasks and that some micro-tasks would not be easy or understandable.

Second, while the students who participated in the study were not professional architects, they were not laypeople either. In general, there is always a challenge to determine whether or not a specific micro-task is appropriate to project stakeholders. To address this concern, for each individual student, we noted the seniority factor so that to establish a relation between each student's experience (in years) and other factors. According to the results, introducing this parameter was practical, as 2^{nd}-year students (i.e., those who only completed one year of training) had, on average, lower performance than their more experienced peers.

Yet, without assuming the students to be professional architects, we reasoned that they would perform reasonably well for the experimental needs. As discussed previously, for several experiments, we hired professional architects from a freelance website. The results of these experiments provided us with essential indications about the participation of real crowd workers and professional architects.

4.3 Tools and Software

The crowdsourcing model was implemented using the JavaScript programming language. The system was based on a 'front-end' program that provides the graphic user interface (GUI) and a 'back-end' program providing a data service using a REST API.

The front-end program was developed using the 'React' framework, a library for building single-page GUI or mobile applications. The various GUI components were written using an extension of the JavaScript language called JSX or JavaScript HTML. Since web browsers cannot execute JSX programs, a production build was generated and served using Express web-server.

Besides 'React,' the front-end program used the following libraries:

- A react component library that uses the Material-Design guidelines[1].
- Axios: A promise-based HTTP client.
- Filepond: A file upload library
- Fingerprintjs2: A library that generates browser fingerprints to identify and log user activity.
- Leaflet: An interactive map library.
- React-router: A library that provides a router mechanism.

1: Material Design is a design language developed by Google in 2014.

- Socket.io: Client for real-time, bi-directional communication with the back-end program.

The 'back-end' was implemented using the Express HTTP server library and used the REST standard to expose numerous API endpoints, such as Tasks, Projects, and Artifacts. We used all the functionality using the middleware design pattern and employed Sequelize, an object-relational mapping library for object-oriented models. The models were stored using a MySQL 7 relational database management system. We also used the following libraries:

- Config: Configuration management.
- Express-sessions: Session management for authentication.
- Winston: A logging library.

Both programs were deployed on a virtual machine running one CPU unit with 2GB RAM and 50GB storage with Ubuntu 16 LTS hosted by Linode in a data center located in London, UK. The system was accessible using the domain 'Architasker.net' allowing various participants to access the system.

After each class, the participants responded to a survey created using 'Google forms.' The results were saved as a spreadsheet.

Furthermore, backups of the application, the database, and artifact files were saved. The backup files were downloaded from the server and analyzed. Upon the analysis of the results, we discussed the findings and decided on changes in subsequent experiments; the software was adapted accordingly. In this way, the planned experiments were adapted to the results.

Results 5

This chapter presents the results of the experiments performed to address the research questions (see Chapter 4). In the subsequent sections, the experiments are presented in chronological order of how they were conducted. The description of each experiment includes its aim, method, generated data, analysis, and conclusions.

The experiments were a part of a workshop for architecture students and were designed as exercises or parts of projects that evolved over several lessons. Each lesson contained several experiments. However, the projects were not developed to full architectural solutions since they served as a platform for the experiments.

A total of 28 experiments were organized as follows. Experiments 1-7 were part of Workshop A, which focused on "Kurpark Kiosk" (Project 1); Experiments 8-15 were part of the "Detached House" (Project 2 and 3); finally, Experiments 16-26 were part of Workshop B that focused on "Idan Tourist Center" (Project 4).

5.1 Kurpark Kiosk (Project 1)

Project 1 aimed to design a small freestanding structure of 300 m^2 (See figure 5.1). The location of Kurpark in Wiesbaden, Germany, was selected due to its familiarity to the present researcher, while the students were not familiar with it. This was done to simulate the typical difference in environment familiarity that characterizes crowdsourcing. Further detail about Kurpark is provided in the project brief below.

5.1.1 Conceptual Sketch Generation Experiment

Experiment 1 explored the primary design task. A brief was provided to the participants who were asked to suggest a solution using a "Conceptual Sketch".

Figure 5.1: Satellite image of Wiesbaden (source: Google maps). Kiosk location is marked by a red polygon

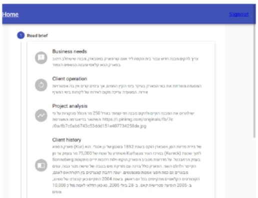


Figure 5.2: Brief screen

Aims

This experiment aimed to 1) provide a proof of concept for the artifact generation task using a 'napkin sketch'; 2) test the hypothesis that a sketch would be sufficient to suggest a preliminary design solution; 3) find out if the task was sufficiently clear and whether valid design solutions were generated; and 4) evaluate the design quality of the generated artifacts.

Method

To perform the experiment, we developed software that implemented the 'preliminary model' (called version 1.0). The following features were provided: 'view task,' 'view brief,' and 'upload sketch.' The database implementation included the following SQL tables: Accounts, Briefs, Sketches, and Uploads. The graphic user interface included the following screens: Brief (Figure 5.2), Task (Figure 5.3), and Upload (Figure 5.4).

The experiment was conducted on November 6, 2018 and was attended by 15 students. The students had different levels of experience: specifically, seven had one year of experience, one had two years of experience, two had with three years of experience, and three students had four years of experience (see Table 5.1). The participants were asked to propose an architectural solution for a kiosk in an urban park in the city of Wiesbaden, Germany. The brief included the following information:

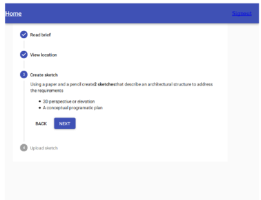


Figure 5.3: Task screen

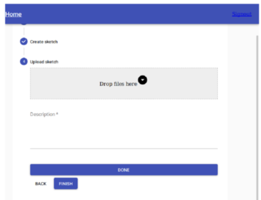


Figure 5.4: Artifact upload

- *Business needs:* A new kiosk has to be built between the Kurpark lake and the ballroom building in Wiesbaden. The building has to fit well into the English-garden design and the adjacent Neo-Classic ballroom building.
- *Client history and setting:* The Kurpark of the Hesse State Capital, Wiesbaden, was established in 1852 in the style of an English garden. It covers an area of 75,000 $m2$ in a narrow valley from the Kurhaus in downtown (Kureck) into the Sonnenberg neighborhood in the Rambach valley. The park is surrounded by large villas from the time of Emperor Wilhelm II. The park features a lake with a 6-foot-high water fountain, old trees, and several monuments. There is an open concert space between the ballroom building and the lake where classical music concerts are held on Sundays.
- *Client operation:* The kiosk serves the park's visitors mainly during the warm summer days. However, on cold days, it does not offer accommodation options. The kiosk needs a new place to host customers in winter.
- *Address:* Sonnenberger Str. 20B, 65193 Wiesbaden, Germany
- *Size:* 300 m^2

The participants had one hour to create a sketch, scan it, and upload it to the application. The participants then filled a survey asking them about their experience. Specifically, the participants were asked what they would improve in the task process and if they felt the need to consult during the work. Finally, the generated artifacts were presented to four professional architects who evaluated the quality of the students' design on a scale from 1 to 5.

Generated Artifacts and Data

The task generated 14 designs and 14 completed survey responses provided by 15 participants. One participant did not complete the task because of technical difficulties (see Figures 5.5 and 5.6).

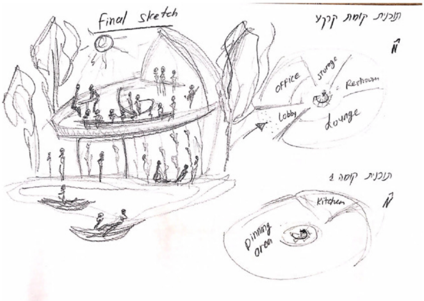
Table 5.1: Generated artifacts, experience, and expert evaluation (R^2 = 0.56) (Experiment 1)

Participant	Experience Years	Artifact	Expert Evaluation
2	3	A13	3
3	1	A21	2
4	1	A26	1
5	4	A11	2
6	3	A4	3
7	1	A12	3
8	4	A3	5
10	2	A30	1
12	1	A7	1
13	1	A19	1
14	1	A20	1
16	4	A8	5
18	1	A29	1
19	1	A9	2

Analysis of Artifacts and Data

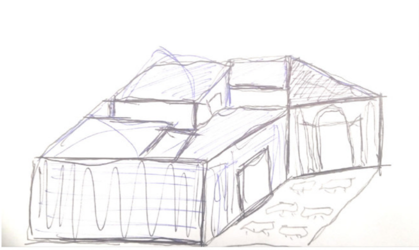
We categorized the generated artifacts taking into account the following aspects:

- *Glassed pavilions* providing open views of the surrounding park and buildings (Artifacts A5, A12, A29, and A30). Artifact A12 had a wave-formed roof.
- *Shell structures* that served as a counterweight to the orchestra shell on the opposite side of the plot (Artifacts A9, A4, and A13). Artifact A13 stood out because of the unique circular floor plan.
- *Neo-classical* styled buildings (Artifacts A6 and A21) that may fit the 19^{th} century design of the park.
- *"L" shaped* structures (Artifacts A20 and A26).
- *Parametric-like* styled structures (Artifacts A3 and A19). Artifact A3 was a parametric-like style wooden

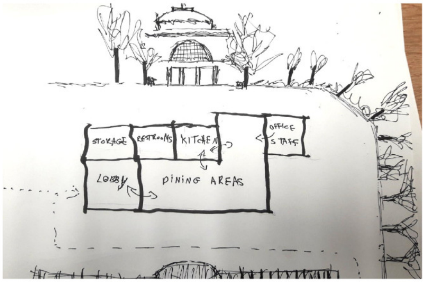


(a) Artifact A13

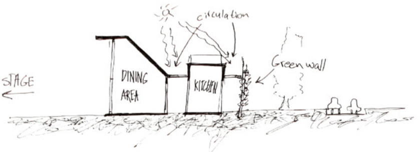
(b) Artifact A21



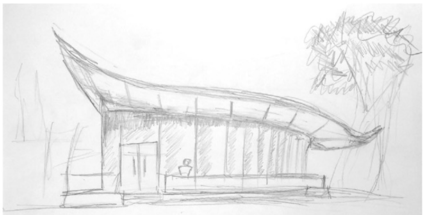
(c) Artifact A26



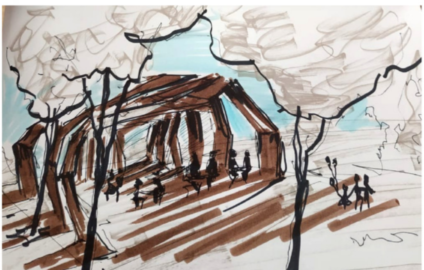
(d) Artifact A11



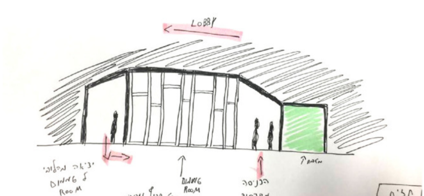
(e) Artifact A4



(f) Artifact A12

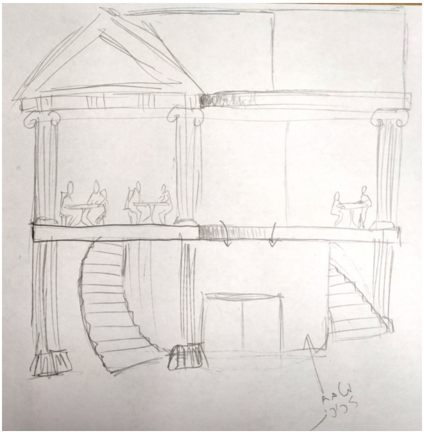


(g) Artifact A3

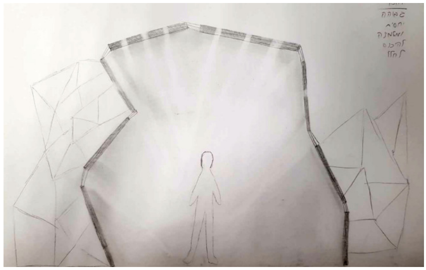


(h) Artifact A30

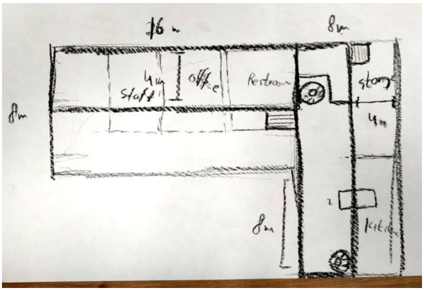
Figure 5.5: Artifact Artifacts (Experiment 1)



(**a**) Artifact A6

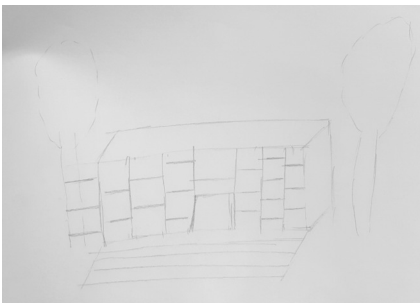


(**b**) Artifact A19



(**c**) Artifact A20

(**d**) Artifact A8



(**e**) Artifact A19

(**f**) Artifact A9

Figure 5.6: Artifact Artifacts (Experiment 1)

frames pavilion. Artifact A19 was a parametric-like styled shell structure.

- *Artificial topography* Artificial topography structure (Artifact A8). The structure was made out of slopes covered by vegetation.

The quality of the sketches varied. Computing Pearson's correlation coefficient between the students' experience and expert evaluation suggested a strong positive correlation between these two variables ($R^2 = 0.56$). Some of the sketches were not successfully scanned and were thus dark, were not appropriately cropped, or did not present the sketch in a good way.

The survey conducted upon task completion showed the participants' positive attitude towards their experience. Of a total of 14 students, 11 participants provided improvement ideas. In eight improvement ideas, the participants mentioned that more time would be needed. Hence, we learned that time was a significant parameter to understand and assimilate the requirements. Furthermore, two participants suggested providing a printed plot map to sketch on. All participants reported that they felt the need to consult with someone about the task, with six answers suggesting communicating with peers or the instructor. There were also requests to provide site plans, sections, and elevations.

Conclusions

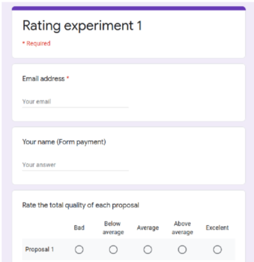
With regard to the design process, the following conclusions were made:

1. The crowdsourced design task was feasible since reasonable-quality artifacts were generated.
2. More time would be necessary to understand the project requirements. We assumed that providing the brief 24 hours before the task would sufficiently prepare the participants for the design task.
3. Sketches are sufficient for a preliminary design solution since most of the artifacts presented an understandable design idea.

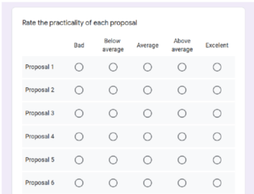
4. The hypothesized positive correlation between expert rating and participants' experience was confirmed by data analysis.

For the software, we concluded the following:

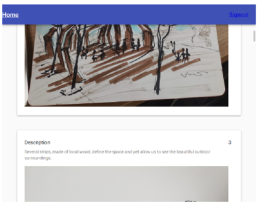
1. The participants liked the design experience using the software.
2. Better scan instructions should be added to the task description.
3. Providing output examples may help students to generate higher-quality artifacts.



(a) Google form



(b) Google form (continued)



(c) Artifact list

Figure 5.7: Screens (Experiment 2)

5.1.2 Artifact Rating Using Scale Experiment

Experiment 2 used a primary artifact-rating task as a proof of concept.

Aims

The experiment aimed to 1) provide proof of concept of the artifact rating task; 2) evaluate the generated data to suggest which artifact should be selected for further development; and 3) learn about the participants' experience rating the artifacts.

Method

The experiment was conducted on the 6th November 2018 with 15 students right after experiment 1 completed. The artifacts generated in experiment 1 were presented with a unique identifier (Screen 5.7c). The participants had to fit the unique identifier to a google-form and provide a rating (Screens 5.7a and 5.7b).

The Google form included the following three sections: 'quality,' 'innovativeness,' and 'practicality.' These 'innovativeness' and 'practicality' metrics are based on Sun et al. (2015) and 'quality' is a general evaluation metric. Each section contained a list of artifact unique-identifiers with a 5-point rating scale [1]. The participants' ratings were

1: 1 = bad, 2 = below average, 3 = average, 4 = above average, and 5 = excellent aggregated, analyzed, and compared to experts' evaluations from Experiment 1. Finally, the participants filled in a survey about their experience; in the survey, they were also asked to provide improvement suggestions.

Generated Data

The participants provided a total of 189 ratings, with the average artifact scores ranging between 2.27 to 3.60 (see Table 5.2). Some ratings were missing from the results (17.7%), meaning that a participant missed some questions. In addition, the ratings for artifacts A29 and A30 were also missing, as the participants submitted these artifacts after the rating task had already started, so they were not included in the form.

Table 5.2: Average artifact rating and standard deviation for each category and expert evaluation (Experiment 2)

Artifact	Total Quality	Innovativeness	Practicality	Average	Expert Evaluation
A3	3.53 (1.06)	3.53 (1.06)	3.6 (1.12)	3.56	5
A4	2.73 (0.59)	2.6 (0.91)	2.8 (0.86)	2.71	3
A7	3.27 (1.16)	3.4 (1.12)	3.29 (0.99)	3.32	1
A8	2.73 (1.1)	2.43 (1.09)	3.2 (1.15)	2.79	5
A9	2.8 (1.08)	2.86 (1.23)	3.13 (1.19)	2.93	2
A11	3.53 (0.92)	3.6 (0.91)	3 (0.93)	3.38	2
A12	2.4 (0.83)	2.27 (0.96)	2.47 (0.99)	2.38	3
A13	2.36 (0.93)	2.27 (0.96)	2.36 (0.93)	2.33	3
A19	2.42 (1)	2.82 (1.4)	2.91 (0.83)	2.71	1
A20	2.77 (1.01)	2.77 (1.09)	3 (1.08)	2.85	1
A21	2.73 (1.19)	2.67 (1.23)	2.92 (1.24)	2.77	2

Data Analysis

An average for each artifact and category was computed. The results showed a high Pearson correlation between 'Quality' and 'Innovativeness' rating ($R^2 = 0.9$). A relatively high correlation was also found between 'Function' with 'Quality' and 'Innovativeness' ($R^2 = 0.65$), suggesting that the participants similarly rated these three categories.

Furthermore, the results of computing the Pearson correlations of the ratings with the expert evaluation showed that the participants and experts disagreed on the chosen artifacts ($R^2 = 0.0025$). However, we found that the participants' ratings were biased: while some students rated their own artifact higher, those same artifacts were rated lower by other students.

The results of the analysis of survey results showed that implementation-wise, the participants perceived the separation of the response form (Google form) from the artifacts list to be inconvenient. First, it required opening two browser windows to load the sketches and a different for the rating form. Second, the form had to be created manually. Some discrepancies found in the ratings were a result of the manual creation of the rating form and the participants' missing some fields.

Conclusions

For the design process, the following conclusions were made:

1. The crowdsourced rating task generated ratings and some agreement among the experts about the quality of artifact A3. However, the task did not result in a selection that correlated with experts' evaluation.
2. If the participants are also the designers, there is a risk of bias in the ratings. Specifically, as shown by our results, the software should not allow participants to rate their own artifacts.
3. While the bias might be responsible for small rating distortion, it should not be critical if there are small grading gaps among the top-rated artifacts. Selecting multiple top-rated artifacts will reduce the effect of small distortions and prevent the disposal of high-potential artifacts.

For the software, the following recommendations can be formulated:

1. The task should be more straightforward and present an image of the artifact next to the rating form.

2. Whenever the task is complex, it should be developed as part of the system, with the rating form being built from the database to prevent gaps in the provided information.

5.1.3 Design Review Experiment

Experiment 3 was a proof of concept and our first attempt to collect design reviews using an online task.

Aims

This experiment aimed to 1) examine the feasibility of the review task; 2) explore how participants provide reviews; and 3) explore how the participants feel about delivering the reviews.

Method

For this experiment, a new review task was developed. The database implementation included the following SQL table: Reviews. The task screen displayed the brief. Next, the review screen showed an artifact, a question with an input field, and all the reviews provided to that artifact (see Figure 5.8). Each time a review was submitted, the page was refreshed, and the next question replaced the current question until all the questions for a specific artifact were answered. After the last answer, the artifact was replaced by the next artifact until reviews were provided to all the artifacts. Other participants' reviews were displayed above the input-form, so participants could see what other participants thought about the artifact.

The experiment was conducted on November 13, 2018, and was attended by ten students. This experiment's input included five top-rated artifacts from Experiment 2 since it was concluded that multiple artifacts should be chosen. The following three questions were presented to the participants for each of the artifacts: 'What did

(a) Task screen displaying reviews

(b) Task screen with input form

Figure 5.8: Screens (Experiment 3)

you like?', 'What did you dislike?' and 'What would you propose to change?' All the questions were answered using a free-text input field without any validation.

Finally, the participants' impressions about the process were collected through a survey.

Generated Data

A total of 118 review items were generated, with between 20 and 28 reviews for each artifact. The average length of a review item was 47.69 characters. The average length of the responses to the 1st question was 62.42 characters, while the average review length of the responses to the other two questions was 36.52 and 40.43 characters long, respectively. Twenty-one reviews were marked as invalid during the analysis of the responses, as the answers were not meaningful. The average percentage of valid responses is reported in Table 5.3.

Finally, seven participants completed the survey. The responses regarding the experience were positive. The participants positively perceived the possibility to see the reviews provided by other participants, as well as the fact that they could provide positive reviews. Some suggested improving the task with the already existing features: seeing other participants' reviews and having a textual artifact description.

Data Analysis

No correlation between the number of reviews provided by a student and the average length of the answers was observed ($R^2 = -0.04$). However, with regard to the review length, the participants produced more extended responses explaining what they liked about the artifact and shorter responses to the question of what they disliked about it or what should be improved.

In the next step, we analyzed and categorized the reviews based on the themes. The 14 identified themes in 4 broad

Question	Responses	Average length	Percentage of valid responses
What did you like?	45	62.42	93.33%
What did you dislike?	34	36.52	67.65%
What would you propose to change?	39	40.43	82.05%

Table 5.3: Reviews distribution by question responses, average length percentage of long responses (Experiment 3)

categories (Program, Form, Environment, Technical Aspects) are presented below.

1. Program:
   a) Circulation - Issues related to the connection of rooms.
   b) Space - Issues related to room sizes.
   c) Function - Issues that prevent proper use of the building.
   d) Program - Issues with the arrangement of program functions.
2. Form:
   a) Concept - Issues related to the design concept.
   b) Shape - Issues related to the shape of the artifact.
   c) Light - Issues related to lighting and natural light.
   d) Roof - Various issues related to roofs.
3. Environment:
   a) Views - Related to things that can be observed from the building.
   b) Scale - Related to scale issues (too small or too large).
   c) Gardening - Issues with trees or fauna.
   d) Surroundings - Issues related to the relation of the artifact with its environment.
4. Technical:
   a) Unclear - The artifact is not sufficiently clear.
   b) Graphics - Issues related to graphic presentation.

As shown in Table 5.4, some artifacts had a very high topic-review ratio, suggesting that multiple and different reviews were provided. However, for some artifacts, a

Table 5.4: Experiment 3: Generated reviews topic analysis

Artifact	Kind	Reviews	Topics	Topics / Reviews
A3	Like	8	7	0.88
	Dislike	9	6	0.67
	Change	6	4	0.67
A8	Like	10	5	0.50
	Dislike	4	1	0.25
	Change	7	5	0.71
A12	Like	10	5	0.50
	Dislike	7	5	0.71
	Change	11	4	0.36
A13	Like	10	9	0.90
	Dislike	8	5	0.63
	Change	8	5	0.63
A30	Like	8	3	0.38
	Dislike	5	2	0.40
	Change	7	3	0.29

relatively low topic-review rate was observed, indicating a higher agreement on the topics between the participants.

We learned that the review task was efficient from the survey, and the participants had a good experience. Some participants noted that they could highlight features they also liked as improvement ideas.

Conclusions

From the results, the following conclusions were made:

1. The crowdsourced review task was feasible and generated review data.
2. A total of 14 review themes emerged from the analysis of the review items. These themes were organized into four categories: Program, Design, Environment, and Technical Aspects.
3. Asking participants what they liked about the artifact may facilitate providing improvement ideas.

5.1.4 Generate 3D Model from Sketch Experiment

Experiment 4 examined the feasibility of the task and the process of generating a model from a sketch.

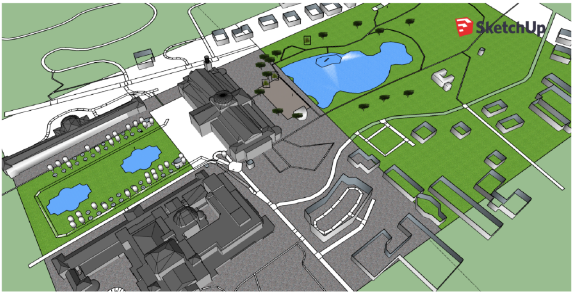
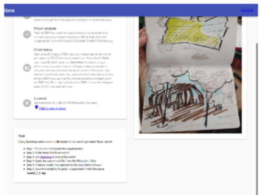


Figure 5.9: Wiesbaden Kurpark - The urban model

(a) Task screen displaying the brief



(b) Task screen with steps

Figure 5.10: Screens (Experiment 4)

Aims

This experiment aimed to 1) provide proof of concept for a 3D model generation task; 2) learn how conceptual sketches are transformed to 3D models, and 3) explore the challenges of using the SketchUp online software for 3D modeling in crowdsourcing.

Method

For this experiment, the software was updated with a new model generation task. First, the brief was presented (Figure 5.10a). The subsequent six task steps (Figure 5.10b) were presented.

The experiment was conducted on November 13, 2018, and was attended by ten students. First, before the participants started the task, a demonstration of how to use the SketchUp software was provided.

After the demonstration, the participants started the task. Each participant was assigned a random sketch out of five sketches selected in Experiment 2.

The first task step was reading the brief and requirements. In the second step, the participants had to download an urban-scale digital model in which they should embed their model. In the third step, the participants had to model the structure. In the fourth step, the participants had to upload the structure and urban model file to Trimble Connect.

The quality of the models was rated by experts. Finally, feedback on the process was collected through a survey.

Generated Designs and Data

Since one participant did not succeed in providing an artifact due to difficulties with using SketchUp, a total of nine new artifacts were generated (see Figures 5.11). Expert architects evaluated the technical and architectural quality of the artifacts (see Table 5.5)

Table 5.5: Generated artifacts, experience, and expert evaluation (R^2 = 0.17) (Experiment 4)

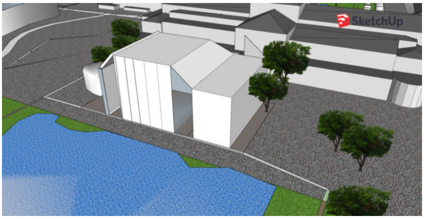
Participant	Experience (Yrs)	Artifact	Expert Evaluation
3	1	A31	2
4	1	A32	1
6	3	A33	4
7	1	A34	2
10	2	A36	1
12	1	A37	2
13	1	A39	2
14	1	A38	3
19	1	A35	3

Finally, seven participants completed the survey. Three participants did not complete the survey because the lesson was over. The answers regarding their experiences varied. Some participants commented that the sketches were not sufficiently detailed, while others noted that they lacked the skills to use the software or that the computers in the lab were not powerful enough for the software.

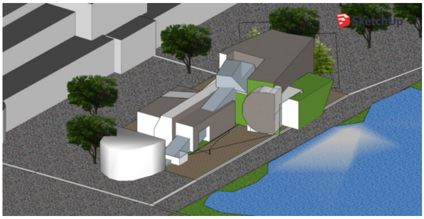
Analysis of Design and Data

In the analysis of the generated artifacts, the following aspects were noted:

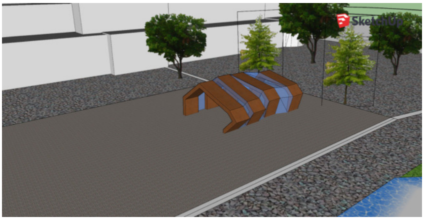
- Artifact A31 and A39 were based on A30 and resembled it in form. A39 was unique since it added black exterior paint to the structure.
- Artifact A32 should have been based on A3 but did not resemble it. We marked it as invalid and omitted it from further analysis.



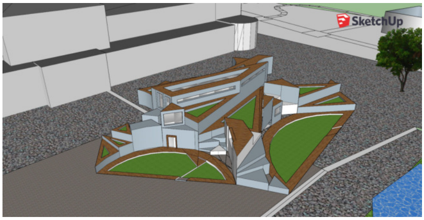
(a) Artifact A31



(b) Artifact A32

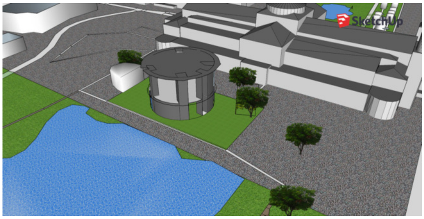


(c) Artifact A33

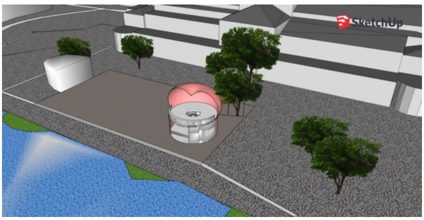


(d) Artifact A34

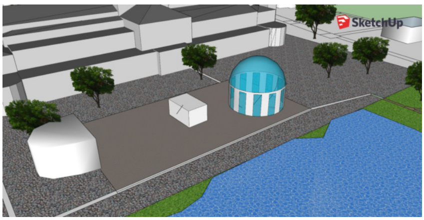
(e) Artifact A35



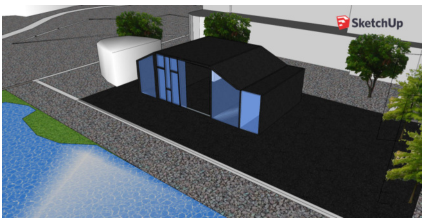
(f) Artifact A36



(g) Artifact A37



(h) Artifact A38



(i) Artifact A39

Figure 5.11: Artifact Artifacts (Experiment 4)

- Artifact A33 was based on A3 and resembled the original artifact while adding glass frames between the wooden frames.
- Artifacts A34 and A35 were based on A8. Both artifacts presented two different interpretations of the original artifact. While A34 was a single stacked structure made of extruded polygons, A35 was made of three multi-story extruded spline structures with connecting bridges.
- Artifacts A36, A37, and A38 were based on A13. While all three were extruded round structures, A37 was the closest adaptation of the original artifact. Artifact A38 was a symmetric domed structure that enclosed the ground floor with walls. Artifact A36 looked unfinished.

The process for converting a sketch into a 3D model is a design process that involves an interpretation to complete information that is not present in the sketch but is essential to the model.

The average expert evaluation of the quality of the models was low (mean. = 2.11). However, there were significant quality differences between the artifacts. While several artifacts were evaluated higher, most were evaluated low or invalid. We also noted a weak correlation between the students' experience and the quality of the artifacts ($R^2 = 0.17$). Nevertheless, A33, which received the highest rating, was created by the most experienced participant.

Before the experiment, we hypothesized that the participants (architecture students) would have the necessary computer skills to use the SketchUp software, especially since it is intuitive. In practice, most participants experienced difficulty in operating unfamiliar software, which degraded their performance.

Conclusions

Based on the results, the following conclusions were made:

1. The crowdsourced model generation task was feasible since we were able to identify a few promising models.
2. There may be a significant knowledge gap when using a design software unfamiliar to the participants, and it is difficult to bridge it in an experimental workshop.
3. Transforming designs from a sketch to the model requires creative interpretation and may result in multiple variations.

5.1.5 Rating Using Categories Experiment

In Experiment 5, we tested the rating and evaluation of artifacts using the jury criteria from the "Safra Square IAUPA competition" described in the pilot experiment chapter.

Aims

The experiment aims were to 1) examine the effectiveness of the rating output for selecting the best artifacts; 2) evaluate the IAUPA competition review criteria; and 3) investigate user experience.

Figure 5.12: Rating task screen (Experiment 5)

Method

A new rating task was developed to replace the previous mechanism used in Experiment 2. The new task screen first presented the brief. Then, a rating screen was displayed. The rating screen included a representation of an artifact, a criteria text, and a rating form (from 1 star to 5 stars, see Figure 5.12). Each time the participant rated a specific artifact, the criteria question changed to a new criteria question. After all ratings for an artifact were provided, the artifact was replaced with the next artifact until all possible ratings were provided.

The "Safra Square" competition evaluation criteria used in Experiment 5 are listed below.

- *Context and continuity:* How does the space flows between the different sections? Do these sections create a hierarchical order?
- *Urban planning:* Does the design provide public welfare? Addressing this question included considerations of protection and use of the environment, as well as the effects on social and economic activities.
- *Connectivity:* Is there a good connection between the design and its urban context?
- *Unique identity:* Does the design provide a unique solution?
- *Composition of masses:* Does the composition of masses create good spaces?
- *Architectural qualities:* Is the design of high aesthetic quality?
- *Simplicity and modesty:* Is the design simple and not too fancy?
- *Sustainability:* Is the design taking the environmental considerations into account?
- *Stability:* Is the design of high structural stability?
- *Flexibility:* How flexible is the design for different uses?

The experiment was conducted on November 20, 2018, with 15 students. It started with the nine artifacts generated in Experiment 4. Upon finishing the task, the participants filled a survey reporting their thoughts about the rating criteria and the user interface.

Generated Data and Designs

The experiment generated 1286 ratings by 15 participates. On average, the participates invested between 5.65 to 12.78 seconds on each rating, and the average was 9.46 seconds. The total time invested by all participants together was 3.4 hours.

Finally, survey responses from the 15 participants were collected.

Table 5.6: Ratings distribution by participant (Experiment 5)

Participant	1 Star	2 Stars	3 Stars	4 Stars	5 Stars	Average Rating (STD)	Rating average time
2	2	15	1	39	33	3.96 (1.11)	8.22
3	3	17	26	32	12	3.37 (1.04)	12.71
4	26	13	15	13	26	3.00 (1.59)	5.65
5	0	10	35	41	7	3.48 (0.78)	12.78
6	20	16	19	27	8	2.86 (1.30)	12.77
7	11	20	32	26	12	3.08 (1.17)	8.83
8	6	6	12	18	48	4.07 (1.24)	8.13
10	2	8	42	31	8	3.38 (0.85)	9.02
12	4	12	25	33	16	3.50 (1.07)	11.06
13	9	29	37	10	5	2.70 (0.98)	10.99
14	18	13	31	17	12	2.91 (1.28)	10.77
16	0	8	55	12	18	3.43 (0.90)	5.39
18	1	0	13	53	23	4.08 (0.70)	9.39
19	6	20	33	19	14	3.16 (1.13)	6.70
Average	7.71	13.36	26.86	26.50	17.29	3.36 (1.08)	9.46

Analysis of Design and Data

With the average rating of 3.36, the distribution resembled a normal distribution where the '3' was the most common, followed by 4, 5, 2, and 1 (see Table 5.6. The average ratings of most participants were around 3. However, these average ratings did not correlate with experts' evaluations. The only exception was Artifact A33 that was evaluated the highest by experts and the second by the participants. Of note, in both evaluations, A33 is one of the two highest-ranked artifacts (along with A35), and the difference in scores of A33 and A35 is very marginal (0.02).

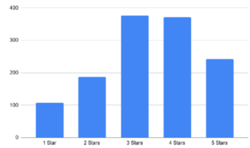


Figure 5.13: Rating distribution (Experiment 5). A normal-like distribution is visible.

While the overall score appeared not to correlate between the experts and participants, some component rating criteria did expose some positive correlation (see Table 5.7). Specifically, in the students' evaluation, the highest correlation criteria were the 'connectivity' criteria ($R^2 = 0.13$), followed by the 'quality' criteria ($R^2 = 0.07$). Both the 'quality' and 'simplicity' ratings provided by the students identified artifact A33 as the highest (as did the experts).

In the follow-up survey, the participants were asked about their experience performing the task, particularly with regard to the specific questions that were asked. One student metaphorically described the questions as "Sisyphic", implying that the questions were difficult to answer. Another student suggested that although the

Table 5.7: Ratings by artifact and rating criteria (Experiment 5)

Artifact	Composition	Connectivity	Context	Flexibility	identity	Quality	Simplicity	Stability	Sustainability	Urban	Average	Expert rating
A31	2.93	3.07	2.93	3.94	2.73	2.69	3.33	4.06	3.20	3.00	3.19	2
A32	3.64	2.92	3.77	3.54	3.85	3.69	3.00	3.54	3.08	3.15	3.42	1
A33	3.40	3.33	3.53	3.93	3.67	*4.33*	*4.07*	3.63	3.73	3.40	3.70	*4*
A34	3.57	3.43	3.31	3.07	*4.07*	3.50	2.79	4.00	3.21	3.00	3.40	2
A35	*4.07*	*3.57*	*4.14*	3.50	3.71	3.50	3.21	*4.21*	3.29	*4.00*	*3.72*	3
A36	3.53	3.27	3.60	*4.00*	3.20	3.67	4.00	4.07	*4.00*	3.60	3.69	1
A37	2.47	2.80	2.93	3.20	2.60	2.67	3.00	3.47	3.20	2.75	2.91	2
A38	2.77	3.00	2.87	2.92	2.92	3.00	3.08	2.92	2.93	2.93	2.93	3
A39	2.93	3.15	2.73	3.54	3.15	3.08	3.69	3.62	3.38	2.77	3.20	2
R^2	0.00	0.13	0.00	0.00	0.00	0.07	0.04	-0.04	0.00	0.03	0.01	

questions were 'good,' the answers were not necessarily positively correlated with what made an artifact best. Yet another student commented that, since the questions were in English (rather than Hebrew), they were more challenging to read and understand.

Regarding user experience, one student suggested displaying the number of the remaining ratings left in the task (i.e., 4 from 120). Another comment was that the new artifacts were loaded too fast, and it was sometimes unclear that the examined artifact was replaced by the next one. Finally, there was a suggestion to add more questions considering additional aspects of the artifact so that to facilitate the correct choice.

Conclusions

Based on the results, the following conclusions were made:

1. The refined rating process worked much better than in Experiment 2 since the artifacts were now presented along with the rating form.
2. The rating displayed a normal distribution, suggesting that the participants provided a balanced rating of the artifacts.
3. Although the distribution was normal, the task failed to identify the best artifact because we were

interested in the best artifacts that were in the 'long tail.' Therefore, while rating on a scale with many categories may provide a good average score, its effectiveness in selecting the best artifact is limited.
4. The quality 'criteria' may have more potential to provide an effective evaluation question for crowd-sourcing.
5. It may be beneficial to select more than one artifact and to allow different idea branches. This would not only allow the development of a variety of ideas but also overcome the concern of inaccuracies in the rating process.

5.1.6 Design Review Experiment

Experiment 6 repeated the review task of Experiment 3, with one artifact only and with an emphasis on the effect of the review data on the outcomes.

Aims

This experiment aimed to explore the effect of displaying reviews by other participants in the review process.

Method

A total of 15 students participated in the experiment on November 20, 2018. The participants reviewed artifact A35 by answering the following three questions: (1) 'What did you like?'; (2) 'What did you not like?' and (3) 'What would you propose to change?'. Unlike in Experiment 3, the participants did not see the reviews provided by other participants.

Finally, the participants filled a survey about their experience.

Generated Data

The 15 participants provided 41 responses. The responses length varied from 4 to 156 characters (mean = 40.04, STD. = 36.35). Further breakdown of the results based on question types can be viewed in 5.8.

Two participants did not answer two out of the three questions due to these students' late arrival to the workshop.

Table 5.8: Generated reviews minimum, maximum, average length (Experiment 6)

Question	Responses	Response Length Min.	Max.	Avg.	Std. Dev.
What did you like?	16	15	156	54.56	42.03
What did you not like?	14	4	84	32.57	28.71
What would you change?	14	4	96	30.92	30.49

Data Analysis

As in Experiment 3, we manually analyzed the reviews and marked invalid responses. The first question had 100% valid responses, while the other two questions had only 64.29% valid responses. These results are similar to the results of Experiment 3 in terms of the average length and validity for each question. The results of a comparison of Experiment 3 and Experiment 6 are summarized in Table 5.9. The results also showed that two participants (4 and 9) provided low-quality responses in terms of text length and validity.

Table 5.9: Comparison of Experiments 6 and 3 by question kind, average text length, and valid response rate

Question	Experiment 6		Experiment 3	
	Avg. length	Valid responses	Avg. length	Valid responses
What did you like?	54.46	100.00%	62.42	93.33%
What did you dislike?	32.57	64.29%	36.52	67.65%
What would you change?	30.92	64.29%	40.43	82.05%

Table 5.10 shows the topics of the review items. As in Experiment 3, the first question continued to receive more diverse responses. We also identify a new review topic, 'Materials'. Most reviews were concerned with the shape (14/16) and circulation (8/16) of the artifact.

Account	Responses	Average length	STD	Valid response rare
2	3	32.33	7.54	100.00%
3	3	51.00	24.75	100.00%
4	3	9.00	4.24	33.33%
5	1	54.00	0.00	100.00%
6	3	58.00	29.63	100.00%
7	1	156.00	0.00	100.00%
8	3	42.00	29.88	100.00%
9	3	13.33	11.81	33.33%
10	3	52.33	50.73	66.67%
12	3	75.67	16.21	100.00%
13	3	31.33	37.95	33.33%
14	3	20.00	11.52	66.67%
16	3	37.33	2.36	100.00%
18	3	20.00	11.78	66.67%
19	3	22.67	13.12	100.00%

Table 5.10: Feedback by participant (Experiment 6)

The survey also asked the students to express their opinions regarding the review process with respect to the questions and the user interface. Some students suggested combining the selection with the review process since these two were very similar, and the students would have liked to express their opinion with rating and text. Another student suggested displaying an interactive 3D model. Some students suggested adding questions, while others thought it would be better if there were fewer questions. Finally, most students were satisfied with the experience, questions, and user interface.

Conclusions

The review task provided multiple feedback items. While some were relevant, others were of low quality. Based on the results, the following conclusions were made:

1. The positive question ('What did you like?') provided longer and more valid responses. Critical questions tended to be shorter and yielded a lower valid response rate. This may also be since it was the first presented question about an artifact.
2. By combining a relative response length response with validity rates, we were able to identify participants who provided low-quality responses.
3. We identified a new review category - 'Materials'.

For the software, the following was concluded:

1. A 3D model should be embedded in the review form.

5.1.7 Model Improvement Experiment

In Experiment 7, we experimented with a task to improve an artifact using the provided feedback.

Aims

The aims of the experiment were to 1) evaluate the feasibility of a design improvement task; 2) identify the challenges of a design improvement process; 3) observe how the participants work with an artifact created by someone else; and 4) learn how the participants would understand the review data and successfully improve the model.

Method

The experiment was conducted on November 19, 2018, with 15 students. The students were provided with one Artifact A45 model and a list of 45 reviews. In order to analyze the effect of the reviews on the outcome, we aggregated the reviews into the following six main themes:

- R1 - Good open and public spaces.
- R2 - The design relates to nature and continues the park.
- R3 - It is nice that the roofs are used for circulation or additional space.
- R4 - Improvement of entrances and openings.
- R5 - Removing the last floor.
- R6 - Buildings should be connected.

At the end of the experiment, the participants filled a survey regarding their experience, and expert architects evaluated the artifacts.

Generated Data and Designs

The participants created 15 new artifacts shown in Figures 5.14 and 5.15. In addition, 15 survey responses were collected.

Analysis of Design and Data

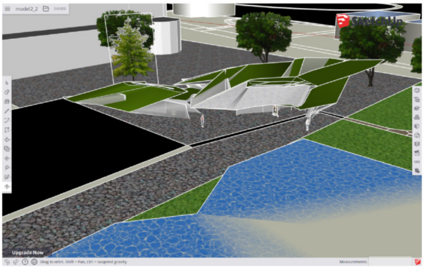
The generated artifacts were analyzed. Four artifacts were identified as 'new design ideas' based on drastic design changes. Seven further artifacts contained various improvements of the original model, improving the relationship between the structures with bridges, adding openings, reducing the structures' size, and adding plants. Four artifacts (A41, A47, A50, A54) were marked as not valid due to their low design quality and random features.

The artifacts that presented a drastic change in the design and contained the following new design ideas:

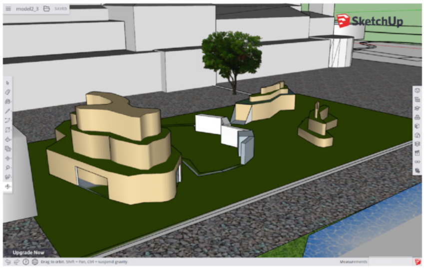
1. Artifact A40: A folded continuous polygon surface structure while diverting from the base structure significantly.
2. Artifact A45: A topographical structure in the form of the number three.
3. Artifact A46: A group of three stacked prismatic structures connected with a bridge. The prisms are dark-colored with some glazed walls.
4. Artifact A51: An extruded polygon with slopes.

The results of the analysis that resolved these issues are presented in Table 5.11. The issue resolve rate was based on the percentage of topics from the resolved reviews. The average rate was 71.21%; the average rate of the new design ideas was slightly lower (62.5%), while the improved artifacts had a slightly higher resolution rate (71.21%).

No correlation between the participants' experience and expert evaluation of their artifacts was found.



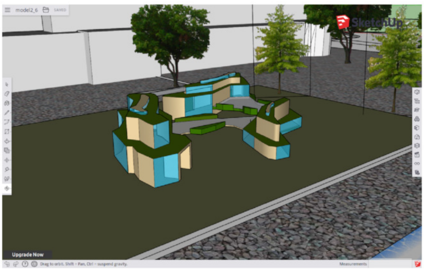
(a) Artifact A40



(b) Artifact A41

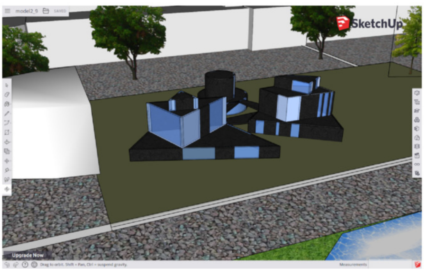
(c) Artifact A42

(d) Artifact A43



(e) Artifact A44

(f) Artifact A45



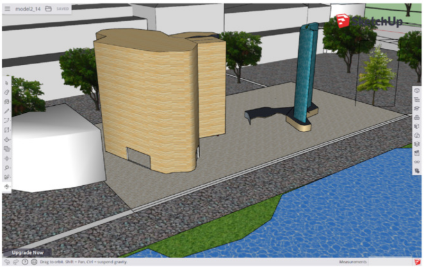
(g) Artifact A46

(h) Artifact A47

Figure 5.14: Artifact Artifacts (Experiment 7)

(a) Artifact A48

(b) Artifact A49



(c) Artifact A50

(d) Artifact A51

(e) Artifact A52

(f) Artifact A53

(g) Artifact A54

Figure 5.15: Artifact Artifacts (Experiment 7)

Table 5.11: Experiment 7 generated artifact expert evaluation, Experience years and review issue analysis

Artifact	Experience Years	Expert Evaluation	Kind	R1	R2	R3	R4	R5	R6	Resolve rate
A40	1	2	New	-	-	-	Yes	Yes	Yes	50%
A45	4	3	New	Yes	Yes	Yes	-	Yes	Yes	33%
A46	1	4	New	-	-	-	Yes	Yes	-	83%
A51	2	1	New	Yes	Yes	Yes	-	Yes	Yes	83%
A42	1	4	Improvement	Yes	Yes	Yes	-	-	-	50%
A43	4	1	Improvement	Yes	Yes	Yes	-	-	-	50%
A44	3	4	Improvement	Yes	Yes	Yes	Yes	Yes	Yes	100%
A48	1	4	Improvement	Yes	Yes	Yes	Yes	-	Yes	83%
A49	1	1	Improvement	Yes	Yes	Yes	Yes	-	Yes	83%
A52	1	3	Improvement	Yes	Yes	Yes	Yes	Yes	Yes	100%
A53	1	1	Improvement	Yes	-	Yes	-	Yes	Yes	66%

In this task, the participants showed a better performance using SketchUp, which may be due to their gaining more experience using this software. We also noticed that the designs were bolder than before.

Some participants found the experience of improving the design enjoyable. However, other participants noted that they experienced difficulties working on a design they did not create and commented that working with SketchUp was still tricky. 47% of the participants thought that they successfully improved the original design.

In contrast, only 13% of the participants thought that their model did not improve the original. In addition, some participants indicated that receiving an existing 3D model was challenging and that allowing them to create new models should be considered.

Conclusions

For the design process, the following conclusions were made:

1. The improvement task was successful, and most artifacts presented an improvement of the original artifact.
2. Four new designs emerged unexpectedly, highlighting a high level of creativity used in this task.
3. 20% of the artifacts were identified as invalid. This ratio is relatively high compared to that observed in Experiment 4, where the participants generated a

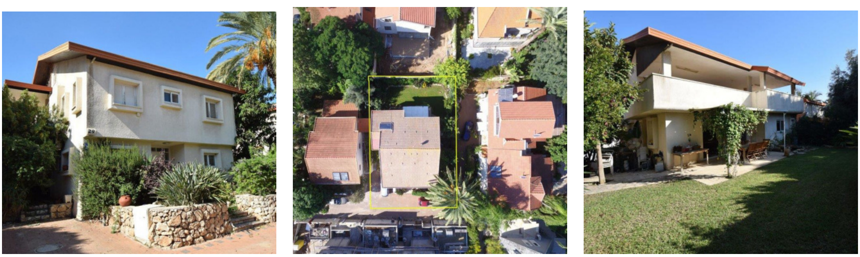
3D model from a sketch. In addition, this result is in contrast to the experience gained using the software. We concluded from the survey indications that this stemmed from the difficulty of handling a given existing model.
4. Most participants paid attention to the review data and improved the model.

5.2 Detached House (Project 2)

Project 2 was to design a new small residential house to replace an existing house in the suburbs of Tel Aviv (see Figure 5.16). The specific plot location was chosen because we had access to the details of the site information, including the surveyor map, photos, and measurement. The project design requirements were to propose an architectural solution for a one-family, 2-story, $200m^2$ building. The building had to include the following components of a standard residential program: Living room, dining room, kitchen with a casual dining place, three bedrooms, bathroom, guest WC, a master bedroom (with a walk-in closet, and bathroom), and a washroom.

5.2.1 Conceptual Sketch Generation Experiment

In Experiment 8, we evaluated whether and, if so, how the improvements made to the 'concept sketch task' would



(a) Street view **(b)** View from above **(c)** Garden view

Figure 5.16: Detached house location

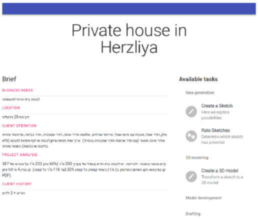


Figure 5.17: Brief screen and task selection (Experiment 8)

Figure 5.18: Sketch task steps (Experiment 8)

Figure 5.19: Sketch task upload (Experiment 8)

improve the concept sketch task output.

Aims

The aims of the experiment were to 1) test whether the enhanced sketch task affected the quality of the design products; and 2) evaluate the participants' satisfaction with the new user experience.

Method

The software was updated to include changes in user experience. A new brief page served as a starting point for all the tasks (Figure 5.17) was developed. The task screen was changed, and the steps were visualized using example images (see Figure 5.18).

The experiment was conducted over two days (November 26-27, 2018) and was attended by 15 students. On the first day of the experiment, 24 hours before the workshop lesson, the participants received an email with the brief. The brief included requirements for a new scheme, including a surveyor map and photos of the site. On the next day, the participants had 1.5 hours to create plans and elevation sketches. Then, they were required to scan and upload the sketches. The students were provided with a printed surveyor map that included a 1x1 meter grid to sketch a plan.

Next, the participants filled a survey about their experience. Finally, the quality of the artifacts was rated by four expert architects. The survey asked the five following questions:

1. Was it easier for you to create a sketch than last time (Experiment 1)? If so, why?
2. Are you satisfied with your sketch?
3. Have you used the surveyor map?
4. Did you get the task details in advance via email, and did you have enough time to think about it?
5. Has the surveyor map contributed or restricted you in your design work?

Generated Designs and Data

The participants generated a total of 15 sketches (B1 - B15). Seven sketches (B8, B10, B11, B12, B13, B14, and B15) were drawn using the printer surveyor map. The sketches included 33 plans, seven elevations, nine perspectives, and two section drawings (see Figure 5.20,5.72, and 5.73). The time difference between the submissions was 53 minutes.

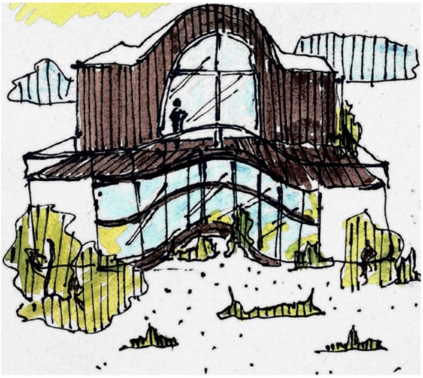
After completing the task, 14 participants answered the survey. One participant did not answer the survey because the workshop lesson was over. Most participants were satisfied with the artifact they made. Ten participants indicated that they used the surveyor map; however, while nine students thought it helped them to create their designs, one student thought it had limited the design. Most participants (10 out of 14) commented that the task was easier than the previous one; the remaining four students thought it was equally or more difficult. Specifically, the following reasons were provided:

- The task was provided in advance, and the participants were ready for it (Process).
- The task was principally easy, or the participants already knew how the software worked (Process).
- The existence of a surveyor map and photos helped to understand the requirements (Process).
- The project requirements were easier / more familiar (Project).
- The project was located in Israel, so the participants were familiar with the culture and local architecture (Project).

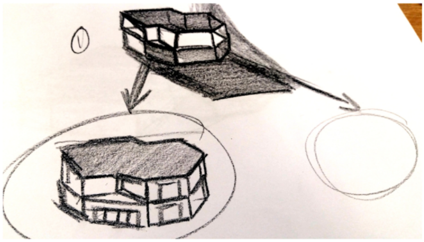
Analysis of Design and Data

All sketches were grouped into the following five categories:

1. 'L' shape (Artifacts B4, B9, and B14). Artifact B4 was an 'L'-shaped house where one wing had a gabled roof with a glassed front.

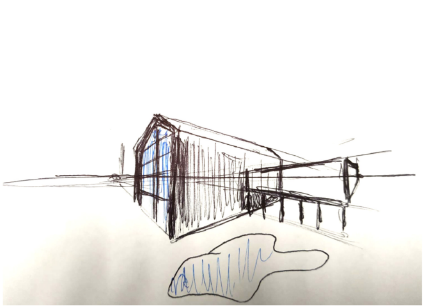


(a) Artifact B1

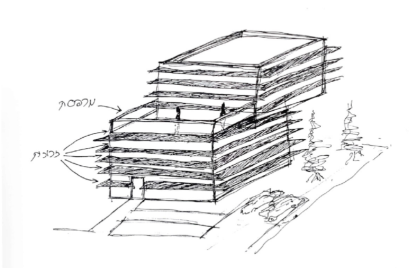


(b) Artifact B2

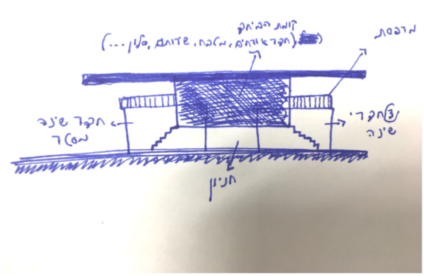
(c) Artifact B3



(d) Artifact B4

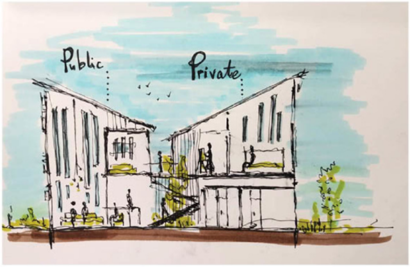


(e) Artifact B5



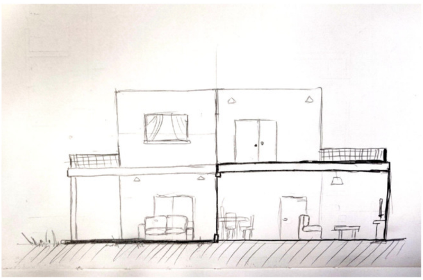
(f) Artifact B6

Figure 5.20: Artifacts (Experiment 8)



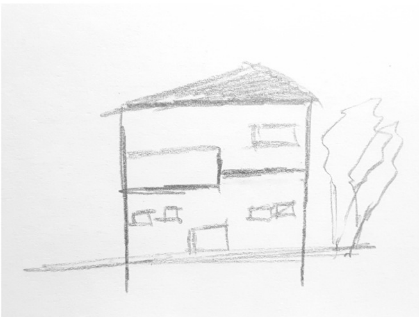
(a) Artifact B7

(b) Artifact B8

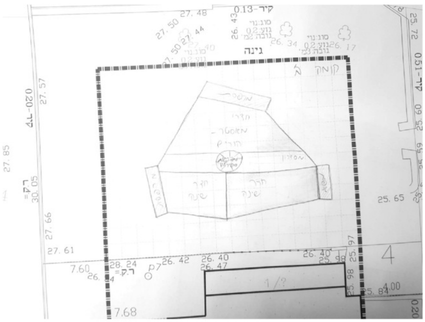


(c) Artifact B9

(d) Artifact B10

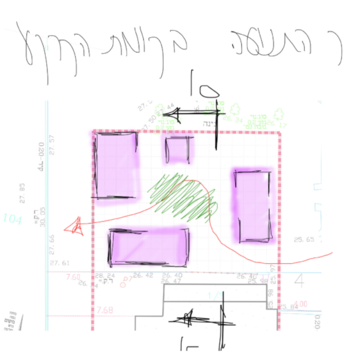


(e) Artifact B11



(f) Artifact B12

Figure 5.21: Artifacts (Experiment 8)



(a) Artifact B13

(b) Artifact B14

(c) Artifact B15

Figure 5.22: Artifacts (Experiment 8)

Table 5.12: Generated designs (Experiment 8). The results showed a correlation between expert evaluation and the participants' experience ($R^2 = 0.67$)

Artifact	Experience (Yrs)	Expert Evaluation	Used map	Plans	Elevations	Perspectives	Sections
B1	4	4	No	1		1	
B2	1	1	No	4		1	
B3	1	1	No	2		1	
B4	1	1	No	4		3	
B5	3	3	No	2		1	
B6	2	2	No	1	1		
B7	4	5	No	1	1	1	
B8	1	1	Yes	2			
B9	1	2	No	2	1		
B10	1	2	Yes	2		1	
B11	1	1	Yes	2	1		
B12	1	1	Yes	3			1
B13	3	2	Yes	4	1		1
B14	1	2	Yes	2	1		
B15	1	3	Yes	1	1		

2. Square shape (Artifacts B1, B3, B8, B11, and B15). Artifact 1 was a two-story box house that embedded a waveform patio. Artifact B3 was a two-story simple rectangular house. Artifact B11 was a simple two-story house with a small patio. Artifact B8 was a two-story rectangle house where the lower story was on columns. Artifact B15 was a two-story rectangular house with a niche entrance and a prominent staircase.
3. Detached buildings (Artifacts B7 and B13). Artifact 7 suggested two sloped boxes buildings: one public and the other private. Artifact B13 suggested distributing the rooms as separated structures on the plot.
4. Stacked extruded polygons (Artifacts B2, B10, and B12).
5. Shifted boxes (Artifacts B5 and B6). Artifact B5 suggested a two-story building made of two boxes where the upper box was shifted to create a balcony. Artifact B6 was an asymmetric multi-story building with a large cantilever roof.

According to the expert impression expressed verbality, there was an improvement in the produced artifacts quality, and most participants reported that they were satisfied with their design outcomes. The participants testified that sending the brief 24 before the task was significant and that the surveyor map and images contributed to a clearer understanding of the task.

The submitted artifacts were made out of multiple drawings—such as plans, sections, elevations, and perspectives - all produced according to each participant's judgment. The average number of drawings provided by each participant was 3.4; however, some artifacts included five, six, or even seven drawings. We noticed that the highest-rated artifacts had no more than three drawings (B1 and B7) and only one drawing of each kind.

The results also showed a high correlation between the participants' experience and the generated artifact (R^2 = 0.67). We could identify the artifacts drafted on the surveyor map. Interestingly, the artifacts were rated by the experts significantly lower than the artifacts that were created freely.

Furthermore, the survey results showed that 71.5% of the participants felt better about the sketching task, and 93% were happy with the sketches they had provided. There may be multiple reasons for these improvements, including providing the brief in advance, the students' experience with the software, and a familiar design project.

Conclusions

Based on the results, the following conclusions were made:

1. Providing more time to process the project brief and requirements in advance was effective.
2. The usage of the printed surveyor map and many photos helped the students to understand the requirements better. However, the surveyor map should not be used as a drafting aid, as the results could be of low quality.
3. Simplifying the task steps, providing clear examples, and familiarity with the software helped the participants perform better.
4. A conceptual sketch should be limited in the numbers of drawings. Allowing participants to submit multiple drawings may result in low quality.

5.2.2 Combined Rate and Review Task Experiment

Experiment 9 tested the feasibility of a combined rate and review task. Since the graphical user interface of both tasks is similar and the participants already invest time in an evaluation of the artifact to provide ratings, they may also provide a text-based review.

Aims

This experiment's aims were to 1) evaluate the combined rating-review task in terms of efficiency, output quality, and participant satisfaction; and 2) compare different evaluation questions in order to find which of them was more related to expert evaluation.

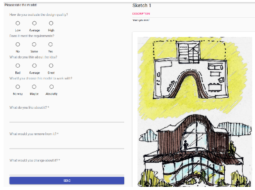


Figure 5.23: Rate and review task screen (Experiment 9)

Method

A new task screen that included a rating and a review form was developed. The new form included the following seven evaluation questions. Questions 1-4 provided answer options; questions 5-7 required a free text response.

1. How do you evaluate design quality? Possible options: 'Low', 'Average', or 'High'.
2. Does it meet the requirements? Possible options: 'No', 'Some' or 'Yes'.
3. What do you think about the idea? Possible options: 'Bad', 'Average', or 'Good'.
4. Would you choose this model to work with? Possible options: 'No way', 'Maybe' or 'Absolutely'.
5. What do you like about it?
6. What would you remove from it?
7. What would you change about it?

First, an artifact was presented along with the review form. Submitting the form refreshed the screen and presented a new artifact for review, until all participants reviewed all possible artifacts. The software did not allow the participants to rate and review the artifacts they produced in the previous experiment to prevent bias. The task screen is shown in Figure 5.23.

The experiment was conducted on November 27, 2018, with 15 students. The participants were required to rate and review 15 artifacts produced in the previous experiment. The task page displayed the artifact sketch files and the description (see Figure 5.23).

Generated Data

A total of 751 ratings and 564 reviews by 15 participants were generated. Some participants did not finish the task after 42 minutes into the experiment, as they came late to class; therefore, the experiment continued until the end without all the participants' finishing their tasks.

The shortest time to conclude the task was 17:28 minutes, while the average was 28:04 minutes (see Table 5.14). The rating results showed that artifacts B1 and B7 received the highest ratings among the participants (see Table 5.13).

Table 5.13: Ratings by artifact (Experiment 9). The ratings show a positive correlation with expert rating ($r^2 = 0.57$)

Artifact	Design Rating	Idea Rating	Requirements Rating	Choose Rating	Expert Evaluation
B1	**93.59%**	**91.03%**	83.33%	**84.62%**	**4**
B2	76.19%	76.19%	78.57%	66.67%	1
B3	59.26%	64.81%	**88.89%**	61.11%	1
B4	60.78%	74.51%	68.63%	62.75%	1
B5	**89.74%**	82.05%	87.18%	**84.62%**	3
B6	56.41%	58.97%	74.36%	56.41%	2
B7	**92.31%**	**89.74%**	**92.31%**	**87.18%**	**5**
B8	45.00%	46.67%	81.67%	48.33%	1
B9	66.67%	66.67%	74.36%	64.10%	2
B10	69.44%	77.78%	**88.89%**	66.67%	2
B11	50.00%	61.11%	75.00%	58.33%	1
B12	55.56%	81.63%	**89.80%**	82.31%	1
B13	68.75%	70.83%	85.42%	69.23%	2
B14	64.29%	66.67%	80.95%	61.90%	2
B15	72.73%	75.76%	75.76%	66.67%	3

To evaluate the quality of the reviews, we counted the number of repetitive text strings and the string length for each participant. We assumed that good feedback would be more extended and less repetitive. The percentage of repetitive reviews is provided in Table 5.14. The results showed that the student who was the fastest to finish the task did not complete the task and provided only 21 feedback items. This was because the student started the task late but had to stop since we could not postpone the next experiment in class. The average time to complete the task (without the aforementioned student) was 28:50, i.e., on average, 1:55 minutes per artifact.

Furthermore, we analyzed the reviews for repetitiveness. The repetition rate for each participant was calculated by counting the non-unique reviews in relation to the total

Table 5.14: Review analysis by participant (Experiment 9)

Account	Reviews	Average Time	Repetitive review rate	Average review length
2	42	0:31:38	19.05%	38.33
3	39	0:40:41	20.51%	50.38
4	42	0:41:15	35.71%	24.43
6	42	0:24:31	69.05%	35.38
7	21	0:17:28	19.05%	23.48
8	42	0:27:09	26.19%	37.12
9	42	0:18:40	64.29%	14.45
10	42	0:22:44	33.33%	36.38
12	42	0:19:39	19.05%	29.38
13	42	0:36:23	9.52%	52.33
14	42	0:29:44	57.14%	15.76
17	42	0:24:50	14.29%	35.36
18	42	0:20:43	57.14%	16.79
19	42	0:37:30	0.00%	58.48

number of reviews. We compare this ratio to the average time it took to provide the reviews and the review text length (see Table 5.14)

Data Analysis

Unlike in previous experiments, the rating distribution in this experiment did not show a normal distribution. There were 419 ratings of '3', 392 ratings of '2', while only 199 ratings of '1' (mean = 2.14). This change may result from the difference in the rating scale that had fewer options.

Three questions showed a high positive correlation - namely, Design quality and Idea ($R^2 = 0.70$), Design quality and Choice ($R^2 = 0.64$); however, the 'requirements' question did not significantly correlated with other questions ($R^2 = 0.129$).

We also compared the expert evaluation with the ratings visible in the graph (see Figure 5.24). There was a strong correlation between 'Design' rating and expert evaluation, while 'the correlation of 'Requirements' with expert evaluation is low. These results may suggest that the expert evaluation was more focused on the quality of the design.

In addition, the high correlation suggests that participants with limited design experience may be able to

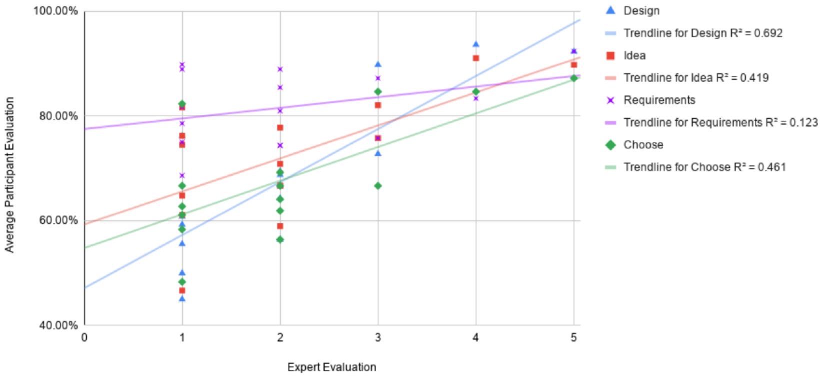


Figure 5.24: Participant rating categories and expert evaluation (Experiment 9)

estimate design quality like professional architects. Therefore, experts may not be required to evaluate and select artifacts.

The total average time to review and rate an artifact was 1:55 minutes, which is rather long for a rating task. However, it is reasonable for a review task. This may be a problem since rating multiple artifacts may require a considerable time investment. Moreover, as most of the artifacts will be removed from the design process, it may be wasteful.

A strongly positive relationship between time and repetition rate was observed when comparing the required time with the average review length and review repetition. The negative correlation between text length and the repetition rate ($R^2 = -0.5$) means that the participants who provided shorter feedback also had a high repetition rate. In contrast, other participants provided more extensive and more unique feedback. Overall, almost most participants displayed some degree of repetition.

Conclusions

Based on the results, the following conclusions were made:

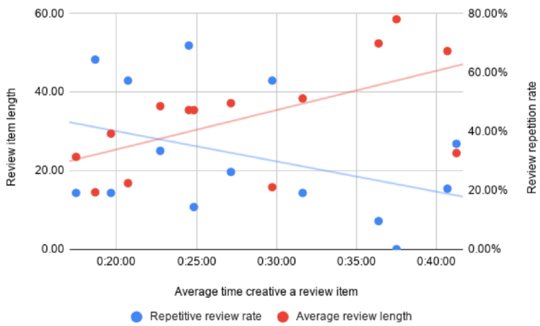


Figure 5.25: Relationship between time, review length, and review repetition (Experiment 9)

1. Feedback tasks and selection tasks should be separated since it is wasteful to provide feedback for discarded artifacts.
2. Providing reviews requires time and should not be overburdened by requiring reviews for many artifacts.
3. 'Design quality' rating was the closest to experts' evaluation.

5.2.3 Architectural Sketch Task Experiment

Experiment 10 was the first experiment, where we tried to develop plans, sections, and elevations artifacts.

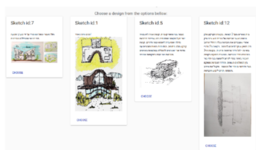


Figure 5.26: Base artifact selection screen (Experiment 10)

Aims

This experiment aimed to learn 1) the performance of a task that requests the transformation of sketch artifact to plans, sections, and elevations from a sketch; 2) how these kinds of artifacts benefit the design process; and 3) if it is possible to have multiple artifacts as the 'base artifact'; the hypothesis was that the selection of the designers would limit the number of idea branches.

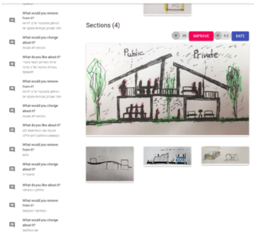


Figure 5.27: Base artifact screen (Experiment 10)

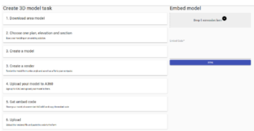


Figure 5.28: Task steps screen (Experiment 10)

Method

The software was updated with a project page created to display the child artifacts of a selected artifact and review information (Figures 5.26, 5.27, and 5.28).

A total of 14 students participated in the experiment that was held on December 4, 2018. In the first part, every participant had to select one of the four provided base-artifacts (B1, B5, B7, and B12). In the next step, the participants were presented with the artifact screen that included the brief, artifact files, and review items. The participants could generate a 'Plan sketch,' 'Elevation sketch,' or 'Section sketch' on that screen. Every participant could submit multiple sketches for different artifacts. The participants were given artistic freedom to create new artifacts.

After the task was completed, a survey about the participants' experience was conducted.

Generated Data and Designs

A total of 39 unique artifacts were generated. Ten artifacts based on B1 are shown in Figures 5.29, 5.30 and 5.31. 11 artifacts based on B5 are shown in Figures 5.32, 5.33 and 5.34. and 16 artifacts based on B7 are shown in Figures 5.35, 5.36,5.37 and 5.38. Finally, two artifacts based on B12 are shown in 5.39.

The distribution of artifact kinds and base-artifact is shown in Table 5.15. Overall, there were 17 plans, 15 elevations, and 11 sections.

With regard to the participants' base-artifact preferences, seven participants chose only one base-artifact, six chose two base-artifacts, and one selected three different base-artifacts. Specifically, B1 was selected six times, B5 seven times, and B7 eight times. Only one participant selected artifact B12 (see Table 5.15).

Expert architects evaluated the artifacts, and the average score for each participant is shown in Table 5.16.

Table 5.15: Artifact distribution by kind and base-artifact (Experiment 10)

Base-artifact	Selecting users	Artifact count	Plans	Elevations	Sections
1	6	9	2	2	5
5	7	11	3	6	2
7	8	17	8	5	4
12	1	2	2	0	0

The expert evaluations positively correlated with the participants' experience ($R^2 = 0.23$).

Table 5.16: Generated artifacts, experience years, and expert evaluation ($R^2 = 0.23$) (Experiment 10)

Account	Artifacts	Experience (Yrs)	Expert Evaluations	Average evaluation
2	B56,B57,B58	3	3,3,3	3
3	B31,B33	1	2,2	2
4	B18,B16	1	1,1	1
5	B30	4	3	3
6	B28,B29,B50	3	4,3,2	3
7	B46,B47,B41,B42	1	1,1,3,1	1.5
8	B55,B48,B43	4	3,3,4	3.3
10	B36,B20,B38,B21,B22	2	3,1,1,2,2	1.8
12	B27,B25	1	2,3	2.5
13	B35,B39,B32,B51	1	2,1,2,2	1.7
14	B45,B54,B49	1	2,3,2	2.3
16	B19,B40	4	1,2	1.5
17	B44	1	3	3
19	B52,B23,B24,B34	1	2,1,1,1	1.2

Finally, a total of 13 survey responses were collected.

Analysis of Design and Data

The task succeeded in creating plans, sections, and elevations from the provided concept sketches. The artifacts provided essential details that made the sketches more explicit.

Most participants selected artifact B7 (8 participants), followed by B5 (7 participants), and B1 (6 participants), while only one participant chose artifact B12. Also, 17 artifacts were developed based on artifact B7, 11 based on B5 and nine based on B1, and only two based on B12. This suggests that artifact B7 was the most popular selection that inspired the creation of more artifacts. This selection correlated with the experts' opinions.

Therefore, it can be assumed that there was a stringent selection process to identify good designs. However, given that the stakeholders are not designers, we needed to implement this method in the selection micro-task.

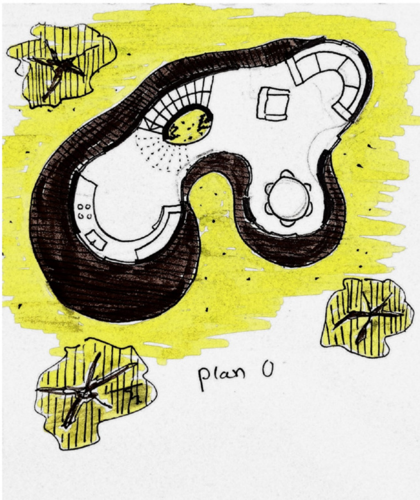
We analyzed the new artifacts as follows:

Six participants created seven new ideas (9 artifacts) from the original artifact B1, which was a two-story box house that embedded a waveform patio.

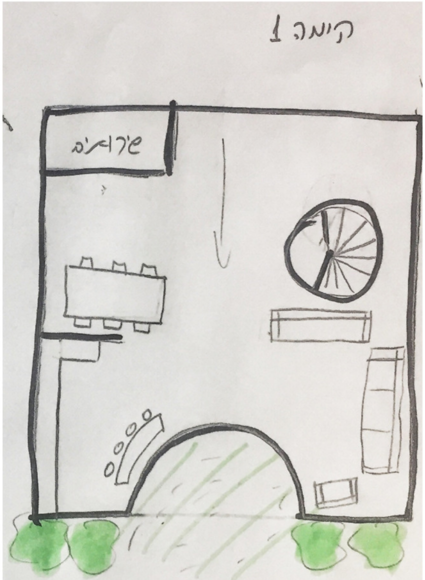
- Artifact B19 intensified the waveform and offered a curved organic floor-plan.
- Artifact B31 was a rectangular plan with a round cut-out similar to the original.
- Artifact B34 was a section very similar to the original.
- Artifact B40 was a section that added pitched roofs and provided opening details and ceiling details.
- Artifact B50 was an adjacent section that showed floating stairs.
- Artifact B55 was a section offering curved walls and roofs.
- Artifacts B56, B57, and B58 were a plan, section, and elevation of a rectangular building without visible similarity to the original. The building had a gable and a sloping roof.

Seven participants created nine new ideas (11 artifacts) based on artifact B5, i.e., a two-story building made out of two boxes where the upper box was shifted in a way to create a balcony.

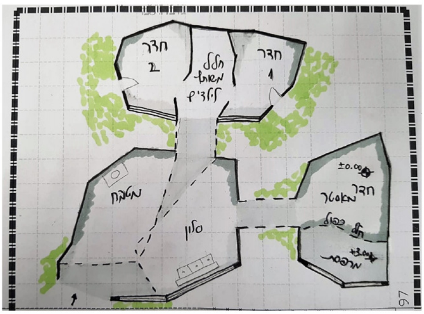
- Artifact B28 was an elevation of the original.
- Artifact B29 was a section of the original.
- Artifact B30 was an elevation with unique triangular openings and with the shades removed.
- Artifacts B36 and B38 suggested that the levels would overlap, creating mid-levels.
- Artifacts B41 and B42 suggested a new elevation composition by fragmenting the two masses on the original rectangular stories.
- Artifact B43 was a plan closely related to the original.
- Artifact B44 was a rectangular plan loosely related to the original.



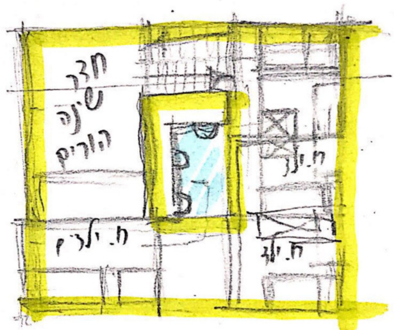
(a) Artifact B19



(b) Artifact B31



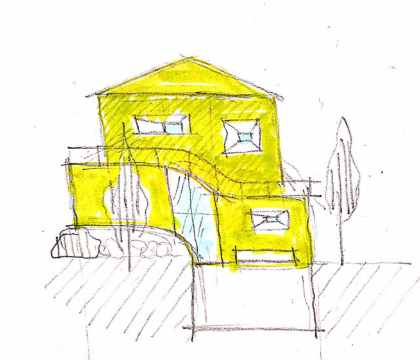
(c) Artifact B32



(d) Artifact B56

Figure 5.29: Plan artifacts based on B1 (Experiment 10)

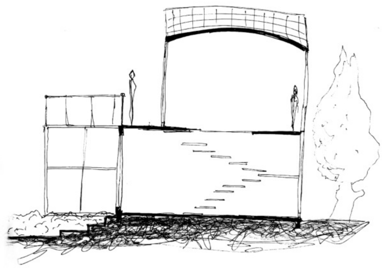
(a) Artifact B34



(b) Artifact B58

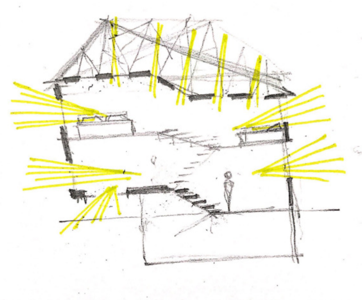
Figure 5.30: Elevation artifacts based on B1 (Experiment 10)

(a) Artifact B40



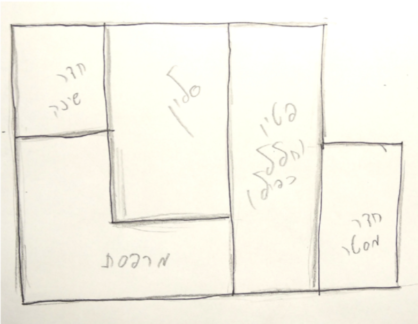
(b) Artifact B50

(c) Artifact B55



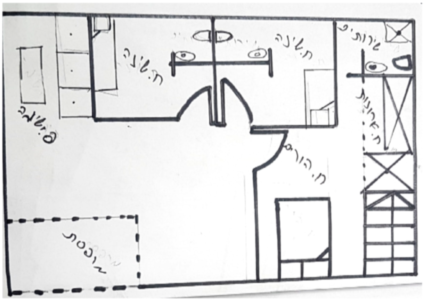
(d) Artifact B57

Figure 5.31: Section artifacts based on B1 (Experiment 10)



(a) Artifact B42

(b) Artifact B43



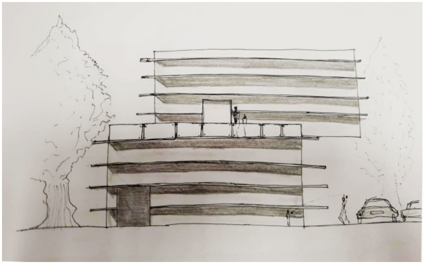
(c) Artifact B44

Figure 5.32: Plan artifacts based on B5 (Experiment 10)

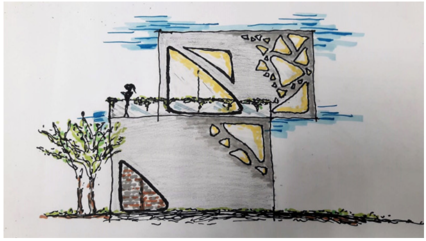
- Artifact B48 was a section of the original.
- Artifact B51 suggested increasing the level shifts of the original to create a four-story building.

Eight participants created ten new ideas (17 artifacts) developed from B7. Artifact B7 was a building made of two wings: one with private rooms and the other with public rooms. The wings were connected via a staircase and a bridge.

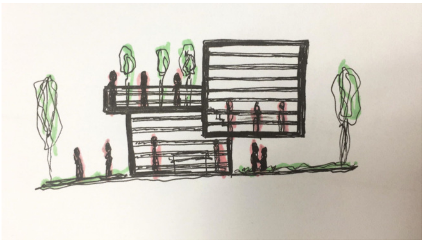
- Artifacts B16 and B18 suggested a rounded opening to the original artifact.
- Artifacts B20, B21, and B22 suggested shifting the second level in each wing similarly to artifact B5.



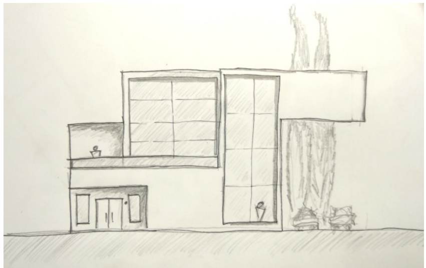
(a) Artifact B28



(b) Artifact B30



(c) Artifact B36

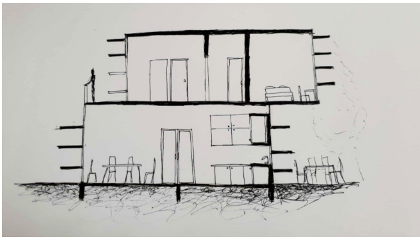


(d) Artifact B41

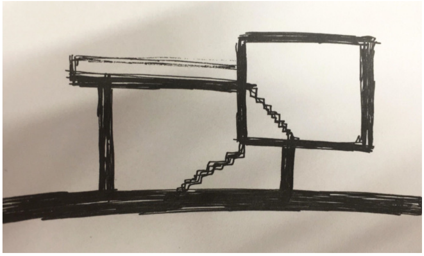
(e) Artifact B48

(f) Artifact B51

Figure 5.33: Elevation artifacts based on B5 (Experiment 10)



(a) Artifact B29



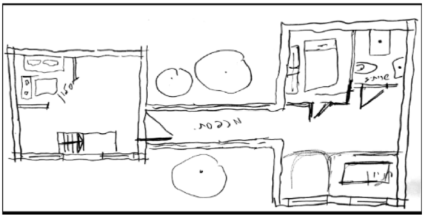
(b) Artifact B38

Figure 5.34: Sections artifacts based on B5 (Experiment 10)

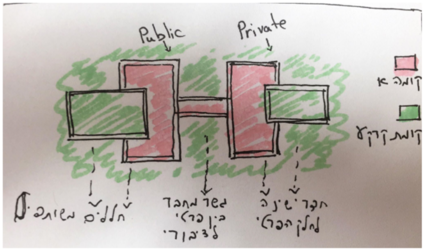
- Artifact B26 was a group of separate prisms. Each prism was a polygon with unique geometry; there was a public outdoor space between the prisms.
- Artifact B27 was a building made out of separated buildings covered with a wave-like roof.
- Artifacts B32, B35, and B39 contained a plan of a building made out of 3 wings connected with open but roofed pathways. The wings were unequilateral polygons.
- Artifact B33 was a plan and elevation of the original but suggested striped windows.
- Artifacts B45 and B54 were plans of the original. One wing had a footprint of an irregular pentagon and the other irregular concaved heptagon. The facades were made out of concrete-like irregular triangles.
- Artifacts B46 and B47 were an ellipsoid building with an irregular glassed facade.
- Artifact B49 was an irregular structure made out of triangular elements.
- Artifact B53 was a two-wing structure with a sloping roof.

Only one participant developed one idea (2 artifacts) from artifact B12, which was a stacked extruded polygon structure.

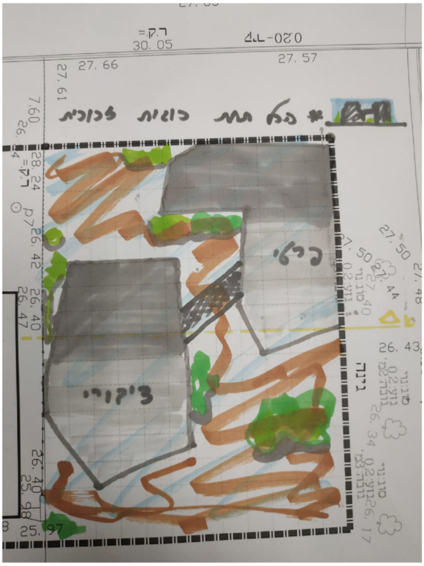
- Artifacts 23 and 24 were polygonal plans made out of parallelograms and a trapezoid with 45-degree angles.



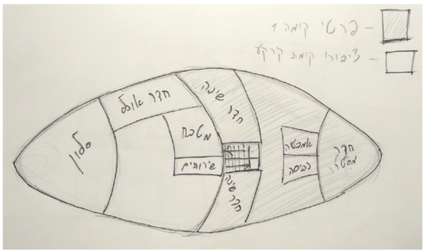
(a) Artifact B16



(b) Artifact B20



(c) Artifact B45

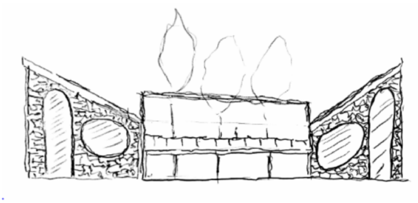


(d) Artifact B46

Figure 5.35: Plan artifacts based on B7 (Experiment 10)

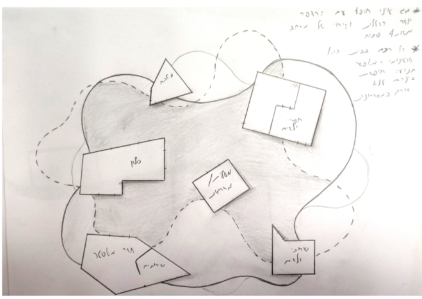
The survey analysis results showed that allowing the participants to select the design they wanted to develop was a pleasant experience. Most participants wrote that the number of base-artifacts was sufficient. One participant commented that there were too many base-artifacts, which limited the number of new artifacts for each base-artifact.

However, 61% of participants reported that they had used the displayed reviews. The reasons included that

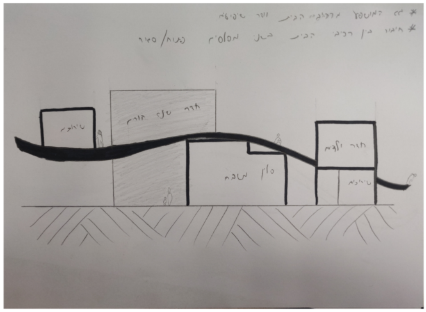


(a) Artifact B18

(b) Artifact B21



(c) Artifact B26



(d) Artifact B27

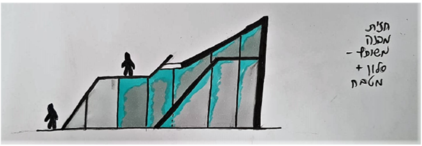
Figure 5.36: Elevation artifacts based on B7 (Experiment 10)

the reviews were not sufficiently highlighted and that they had a high level of repetition.

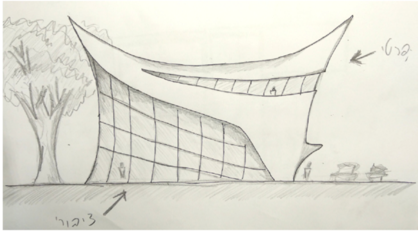
Conclusions

The following conclusions were drawn based on the results:

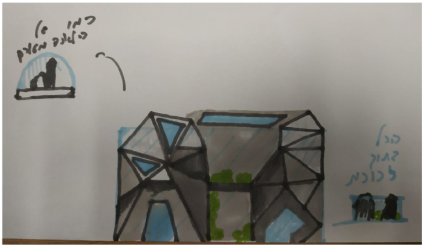
1. The task was successful in creating valuable artifacts: plans, sections, and elevations.
2. The selection of an artifact from an array of artifacts, based on the designer preference, is a suitable selection method to identify design quality.
3. However, it is not very efficient for designers to work on artifacts that may later be discarded. The optimal number of artifacts to provide in this task should be determined.



(a) Artifact B35

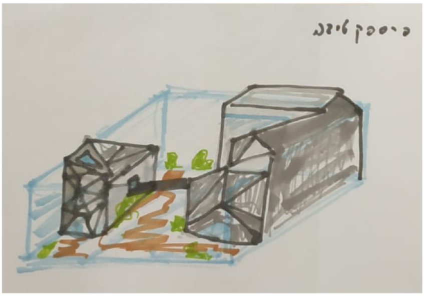


(b) Artifact B47



(c) Artifact B49

(d) Artifact B53



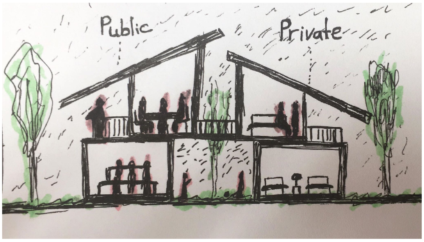
(e) Artifact B54

Figure 5.37: Elevation artifacts based on B7 (Experiment 10)

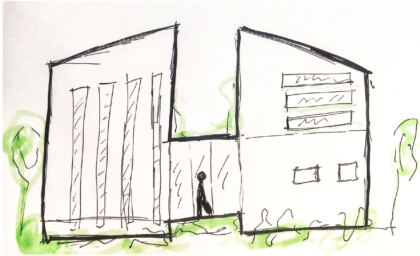
4. Some participants oversaw the reviews. Therefore, reviews need to be displayed in a useful way.

5.2.4 Selection by Comparison Task

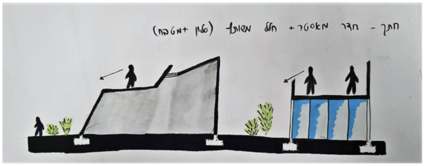
Experiment 11 explored a simple rating method based on a comparison between two artifacts. The idea behind the new rating mechanism was to reduce the cognitive effort in the rating process.



(a) Artifact B22

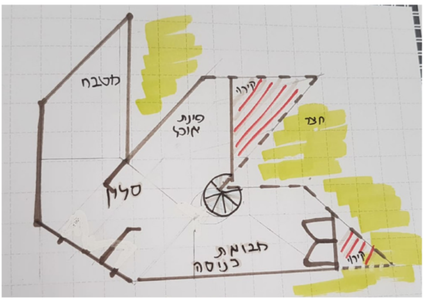


(b) Artifact B33

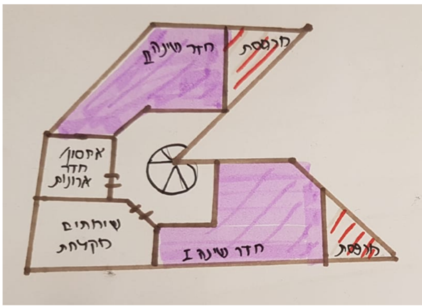


(c) Artifact B39

Figure 5.38: Section artifacts based on B7 (Experiment 10)



(a) Artifact B23



(b) Artifact B24

Figure 5.39: Plan artifacts based on B12 (Experiment 10)

Aims

This experiment aimed to 1) test a new graphical user interface for rating artifacts; and 2) learn how the output data can be used to select the best artifacts.

Method

The software was updated to include a new rating task. After selecting the task, the participants were shown a modified rating screen with two artifacts (Figure 5.40). Above the artifacts, the following question was shown: "Which design is better?". The button with the text "This one!" was placed below each artifact. Clicking on the button created a positive rating record for the selected artifact; when the artifact was not selected, a negative rating record was created.

Figure 5.40: Selection task screen (Experiment 11)

We hypothesized that calculating the average of all responses would yield a better understanding of the best artifact since the best artifacts would receive a positive rating when presented with other artifacts. We restricted the participants from rating the artifacts that they had created.

The experiment was conducted right after Experiment 10 on December 4, 2018, with 14 students. After the experiment, the participants answered a short survey about their experience.

Generated Data

Within 1:05 hours, a total of 851 ratings for 37 artifacts were generated. The experiment finished when the workshop class was over, with only partial task completion.

Table 5.18 shows that the average rating took, on average, 29 seconds. Since the task started after the previous design task, some participants provided various working times. On average, the participants provided 35.86 ratings; therefore, each participant produced 17.93 ratings (since every rating task created two ratings).

Data Analysis

The results showed that the graphical user interface was intuitive. The participants indicated that the rating task was easy, since a comparison between two artifacts was

Table 5.17: Participant and expert ratings (Experiment 11)

Artifact	Rate Count	Positive Count	Negative Count	Rating Sum	Rating Avg.	Expert rating
16	48	48	0	48	100.00%	1.00
18	30	25	5	20	66.67%	1.00
19	3	3	0	3	100.00%	1.00
20	47	41	6	35	74.47%	2.00
21	24	14	10	4	16.67%	2.00
22	14	14	0	14	100.00%	2.00
23	8	8	0	8	100.00%	1.00
24	8	0	8	-8	-100.00%	1.00
25	49	36	13	23	46.94%	3.00
27	13	8	5	3	23.08%	2.00
28	30	30	0	30	100.00%	4.00
29	4	4	0	4	100.00%	3.00
30	31	25	6	19	61.29%	3.00
32	48	23	25	-2	-4.17%	2.00
33	46	16	30	-14	-30.43%	2.00
34	13	13	0	13	100.00%	1.00
35	25	12	13	-1	-4.00%	2.00
36	31	19	12	7	22.58%	3.00
38	4	0	4	-4	-100.00%	1.00
39	14	3	11	-8	-57.14%	1.00
40	14	7	7	0	0.00%	2.00
41	30	12	18	-6	-20.00%	3.00
42	27	18	9	9	33.33%	1.00
43	26	8	18	-10	-38.46%	4.00
44	26	0	26	-26	-100.00%	3.00
45	46	10	36	-26	-56.52%	2.00
46	46	5	41	-36	-78.26%	1.00
47	26	9	17	-8	-30.77%	1.00
48	34	6	28	-22	-64.71%	3.00
49	28	3	25	-22	-78.57%	2.00
50	12	3	9	-6	-50.00%	2.00
51	28	0	28	-28	-100.00%	2.00
52	17	0	17	-17	-100.00%	1.00
55	6	0	6	-6	-100.00%	3.00
56	3	0	3	-3	-100.00%	3.00
57	1	0	1	-1	-100.00%	3.00
58	1	0	1	-1	-100.00%	3.00

Account	Rating count	Total time	Avg. time
2	4	0:02:38	0:00:40
3	83	0:20:11	0:00:15
4	71	0:26:05	0:00:22
5	28	0:18:45	0:00:40
6	48	0:22:32	0:00:28
7	43	0:09:18	0:00:13
8	1	0:00:01	0:00:01
10	48	0:10:03	0:00:13
12	72	0:22:46	0:00:19
13	12	0:19:20	0:01:37
14	24	0:03:57	0:00:10
16	37	0:07:18	0:00:12
17	13	0:07:30	0:00:35
19	18	0:18:03	0:01:00
Average	35.86	0:17:18	0:00:29

Table 5.18: Participant ratings count and time (Experiment 11)

effortless. We learned that selecting a better artifact was more natural than deciding on the rating.

The average ratings did not indicate the best artifacts. The rating did not correlate with the expert evaluation (see Figure 5.41). Computing the Pearson correlation coefficient resulted in a negligible result ($r^2 < -0.01$). An additional check showed that there was no issue with experts' evaluation, as their ratings were reasonable.

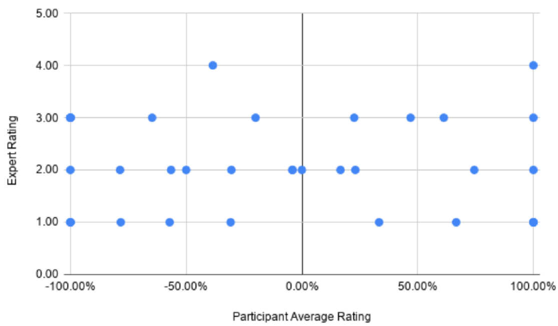


Figure 5.41: A comparison of average participants' ratings and expert evaluation (Experiment 11)

Therefore, this discrepancy could result from the partial ratings, as the experiment finished when the class ended.

Although the task was not successful in identifying the best artifacts, 77% of participants confirmed that the rating user-interface was better than in the previous selection experiment (Experiment 9), and they found it easier to rate artifacts using the comparison. In addition, it was simpler to choose an artifact instead of providing a rating from 1 to 5.

Conclusions

The results suggested the following conclusions:

1. A comparison-based user interface was favored by the participants and was simpler to use.
2. Artifact selection is user-friendlier than rating using a scale.

3. The task failed to identify the best artifacts, as the task was incomplete.
4. The selection system required too much time, and the results were distorted. Other methods can be expected to provide simpler and more robust results.

5.2.5 Merge Sketches to Model Task Experiment

Experiment 12 was the second time we experimented with a task that generated 3D models from sketches. In this experiment, the participants were asked to choose several artifacts and 'merge' them into one artifact.

Aims

The aims of this experiment were to 1) test the production of a 3D model by merging different sketches; and 2) evaluate a new workflow that would allow for displaying interactive 3D models using the 'Autodesk A360' cloud service instead of 'Trimble Connect.' Using Trimble in the previous experiments did not have the option of embedding interactive 3D models on our website. This is necessary for the review and selection tasks.

Method

The experiment was conducted on December 11, 2018, with 14 students. At this stage, we had four branches based on artifacts B1, B5, B7, and B12. Since artifacts B5 and B7 had more sub-artifacts, we decided to remove artifacts B1 and B12. Each of the artifacts had multiple plans, elevations, and section sketches. For the development of a 3D model, the participants had to select one of the branches (B5 or B7) and choose one plan, one section, and one elevation sketch.

Upon completion of the model creation, the students had to render an image of the model and upload it to the website. Then they were required to upload the model

Table 5.19: Generated artifacts, experience, and expert evaluation (Experiment 12)

Artifact	Experience (Yrs)	Expert Evaluation
B59	4	2.5
B60	1	3.5
B61	3	3
B62	3	2.5
B63	1	3
B64	4	4
B65	1	3
B66	1	3.5
B67	4	2.5
B68	1	3
B69	1	-
B70	1	-
B71	1	1.5
B72	2	2.5

to 'A360', a website by Autodesk that allows for sharing 3D files using an online 3D viewer. The participants had to copy the 'embed code' from A360 and input it to our website. After the experiment was completed, expert architects reviewed the generated artifacts.

Generated Data and Designs

A total of 14 task submissions were recorded, but only 12 were completed and provided a rendered image and a model (see Figure 5.42, 5.43, and 5.44). The artifacts are listed in Table 5.19 along with the participants' experience years and expert evaluation.

Analysis of Design and Data

Six artifacts were developed from artifact B5, and eight artifacts were developed from artifact B7. However, artifacts B69 and B70 were lacking the model embed code and, therefore, were invalid.

The artifacts developed from B5 included the following:

- B61 was a two-story glass glazed structure, where the 2^{nd} story was shifted. Horizontal wooden shades covered the glass facades.
- B63 and B67 were two-story structures made out of two interlacing boxes. Diagonal openings perforated the boxes.

- B66 was a three-story modern style structure made out of a combination of vertical and horizontal elements, such as strip windows and multi-story curtain walls.
- B72 was a two-story structure made out of two shifted boxes. It included a colonnade that supported the standing out of the second story.

The artifacts generated from B7 included the following:

- B59 was a two-ring structure made out of folded polygons.
- B60 was made out of two separate and overlapping structures. One structure was two stories high with the second floor protruding. The second structure was three stories high. The third story protruded above the first structure. An external spiral staircase connected the two structures.
- B62 was made out of two continuous folded polygonal surface structures made out of wooden panels.
- B64 was a building made of two wings, both with large sloped roofs.
- B65 was a complex of three extruded polygonal prism buildings.
- B68 was two polygonal extruded prism buildings connected with a bridge on the second floor. The exterior was made out of concrete and triangular windows.
- B71 was a complex of one-story rectangular buildings connected by bridges.

There was a high correlation between the participants' experience and expert evaluation. While the top-rated artifact (B64) was created by a senior participant, the invalid artifacts were made by junior participants (B69 and B70). However, since some junior participants produced high-quality artifacts (B60, B65), the correlation between experience and artifact quality was not strong.

Conclusions

Based on the results of this experiment, the following conclusions were made:

1. Merging multiple sketch artifacts to a single model was a feasible task and produced reasonable artifacts.
2. Using 'A360 'allowed for embedding the 3D model in the application but required more work and complexity, which resulted in two invalid submissions.

5.2.6 Model Improve Task Experiment

Experiment 13 sought to improve 3D-model artifacts as part of an evolutionary improvement process.

Aims

The aims of this experiment were to 1) evaluate the feasibility of a model improvement task based on reviews; and 2) assess whether the 'free selection' of artifacts by the participants was a feasible rating process.

Method

The experiment was conducted on December 18, 2018, with the participation of 11 students.

In the first stage, the participants were requested to generate reviews. They were provided with artifacts from B59 to B72. In the next step, the participants selected one of the artifacts and downloaded their files. Then, they had 40 minutes to read the provided reviews, to improve one issue and to upload the new artifact. These artifacts were called generation 1.

In the next step, the participants reviewed generation 1 artifacts, downloaded one artifact file, improved one issue, and uploaded the new artifact. The second set of artifacts was called generation 2. When done, the participants filled in a survey with questions about their experience in this experiment.

Finally, the artifacts were scored by expert architects to generate a qualitative evaluation of the design.

Generated Data and Designs

In the first stage, a total of 290 review items were generated by 11 participants for the 11 new artifacts. In the next stage, the participants selected five out of the 11 artifacts. Artifacts B60, B61, B63, and B64 were selected twice, while artifact B66 was selected three times. Therefore, nine artifacts were not selected.

The 11 new artifacts generated in the first stage are described below (see Figures 5.45, 5.46, 5.47, 5.48, and 5.49).

Artifact	Experience (Yrs)	Expert Evaluation	Selections
B59	4	2.5	0
B60	1	3.5	2
B61	3	3	2
B62	3	2.5	0
B63	1	3	2
B64	4	4	2
B65	1	3	0
B66	1	3.5	3
B67	4	2.5	0
B68	1	3	0
B69	1	2	0
B70	1	1.5	0
B71	1	1.5	0
B72	2	2.5	0

Table 5.20: Generation 1: Participants' experience, expert evaluation, and number of selections (Experiment 13)

- Artifacts B73 and B77 were based on B60. B73 looked identical to B60. B77 improved the facade of the building and changed some material aspects.
- Artifacts B79 and B84 were based on artifact B61. B79 introduced a sloping new tiled roof. B84 removed the surrounding shades and suggested transforming the structure into a patio building without windows. However, this artifact was marked as invalid, since it was submitted without a model file.

- Artifact B80 and B82 were based on artifact B63. B80 was similar to the original, with some of the facades distorted. B82 added transparent extruded prisms.
- Artifact B74 and B76 were based on artifact B64. B74 and B76 were identical to B64, and B76 had materials replaced.
- Artifacts B78, B81, and B86 were based on B66. B78 had a triangular 3D pattern improvement. B86 was almost identical to the original, with slight material changes. B81 also had some minor changes, mainly removing the window tiling.

In the next stage, the participants generated another 11 artifacts. These artifacts are described below (see Figures 5.50 and 5.51).

Table 5.21: Generation 2: Participants' experience and expert evaluation (Experiment 13)

Artifact	Experience (Yrs)	Expert Evaluation
B87	2	1
B88	3	1
B89	5	3
B90	5	1
B91	4	2
B92	5	2
B93	2	2
B94	4	3
B95	2	2
B96	2	2
B97	2	1

- B89, B94, B95, and B97 were based on B77. B89, B94, B95 changed mainly the windows and divided them. B97 was a new design with a hinged floor over a pedestal.
- B91 was based on B80 and changed the facades, providing new triangular windows.
- B92 was based on B65 and added bridges between different parts of the building.
- B96 was based on B68, removing the openings and painting it in black.
- B88 and B93 were based on B82. B88 and B93 were messy compositions of extruded prisms, so these artifacts were considered invalid.
- B87 was based on B66. The new design looked as if the designer just extruded the building walls, so this artifact was also discarded.

Finally, ten survey responses were collected, and the experts reviewed and scored the generated artifacts.

Analysis of Design and Data

According to expert evaluation, there was an average improvement in Generation 1 artifacts (Figure 5.52). The maximum expert rating was considered to be a success factor for the artifacts' improvement since we expected the rating to increase. The maximum artifact evaluation score between Generations 0 and 1 decreased. Also, both artifact average and maximum scores decreased in Generation 2.

In the survey, we asked the participants to rate on a scale from 1 to 5, with an average of 3, if they had successfully improved their model. We also asked them to rate if the time was sufficient for the task; The analysis of survey results showed that 50% of the participants found the 'improving models that others created' 'interesting' and 'fun,' while 20% of the participants reported that it is was 'hard' and 'not simple.'

The decline in quality and the survey results may suggest that improving artifacts may be a complicated task that requires more time and a high level of expertise.

As concerns the 'free selection,' the results showed that the participants selected artifacts with a higher expert evaluation. Regression analysis showed a positive correlation ($R^2 = 0.52$) between the expert evaluation and the number of participants who selected that artifact (Figure 5.53). This selection was performed by the designers, with an emphasis that, in this kind of process, the stakeholders do not participate.

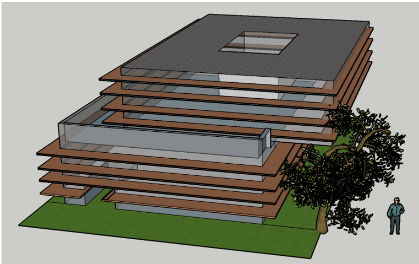
The positive correlation between the expert rating and participant artifact selection in Generation 1 ($R^2 = 0.52$) was consistent with the results of previous experiments. There were some similarities between the selections and expert scores in Generation 2. However, none of the participants selected the top four artifacts by expert rating. This outcome could be related to some bias of the participants, as they may have relied on other criteria when selecting the artifacts. On the other hand, the experts might not have liked the novelty and uniqueness of artifact B77, although it was an exciting and challenging design.

Furthermore, no evidence was found that would suggest that the designers used the reviews to improve the models.

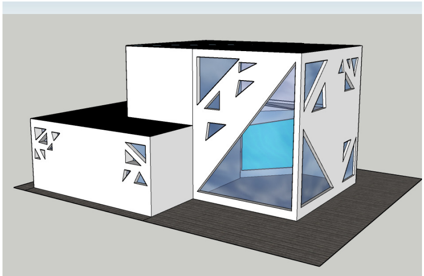
Conclusions

The results suggest the following conclusions:

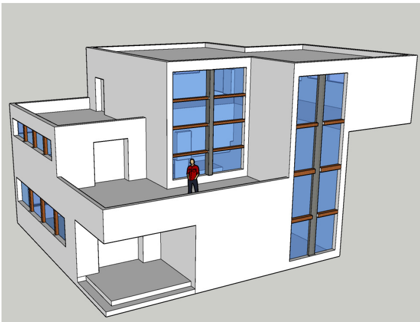
1. The task was unsuccessful in improving the provided model. This may be related to the limited time for the design task.
2. There was a positive relationship between free artifact selection and expert evaluation.
3. Upon generating several selections, a designer may be biased and select designs that are not the best.



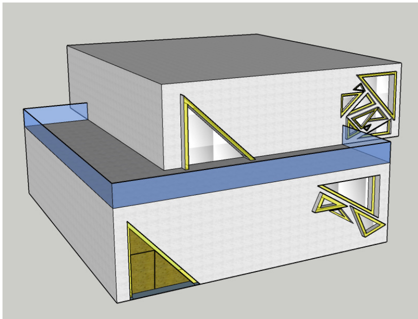
(a) Artifact B61



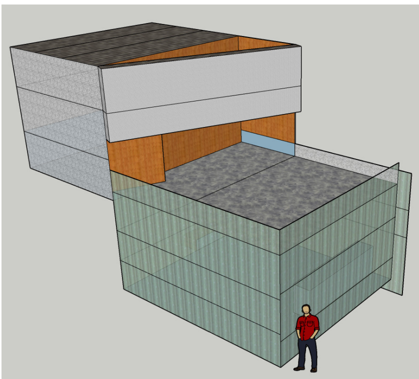
(b) Artifact B63



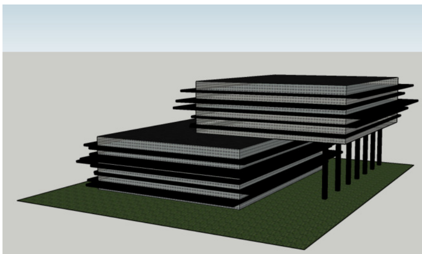
(c) Artifact B66



(d) Artifact B67



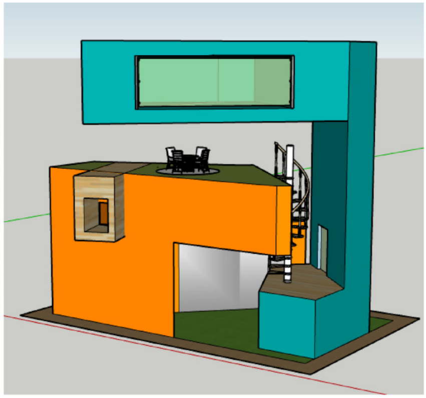
(e) Artifact B69



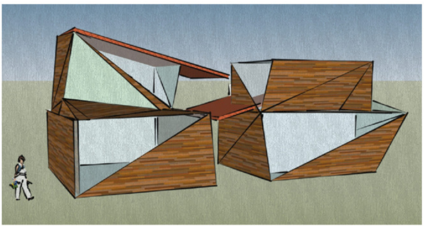
(f) Artifact B72

Figure 5.42: Artifacts based on B5 (Experiment 12)

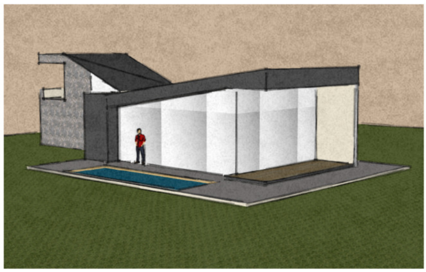
(a) Artifact B59



(b) Artifact B60

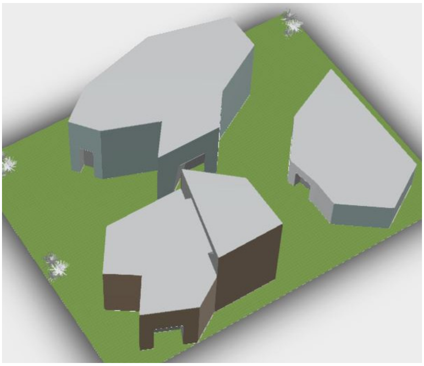


(c) Artifact B62

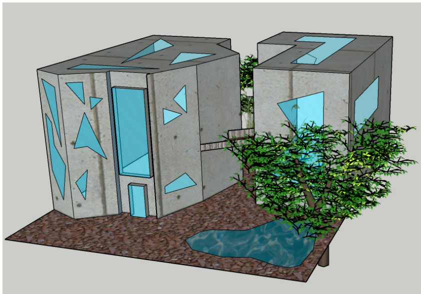


(d) Artifact B64

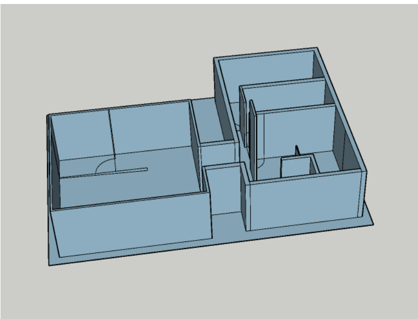
Figure 5.43: Artifacts based on B7 (Experiment 12)



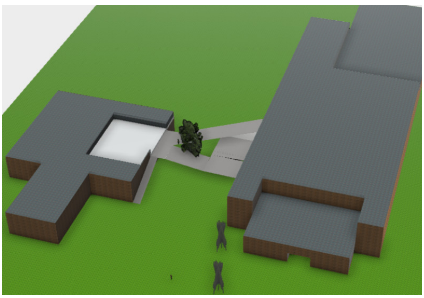
(a) Artifact B65



(b) Artifact B68

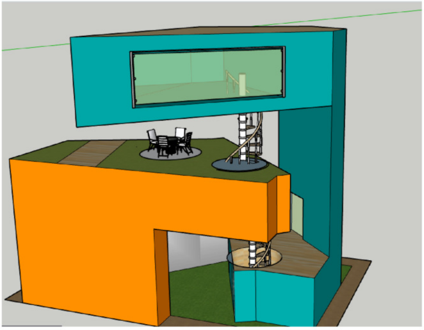


(c) Artifact B70

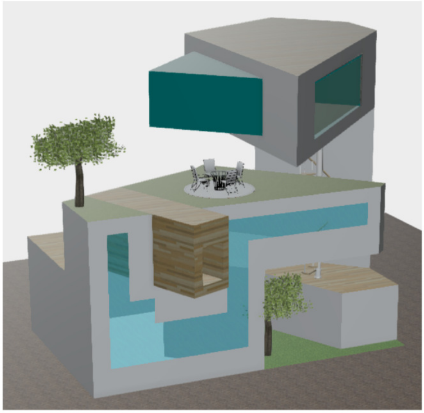


(d) Artifact B71

Figure 5.44: Artifacts based on B7 (Experiment 12)

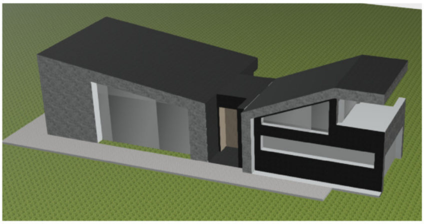


(a) Artifact B73

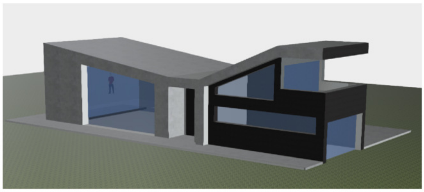


(b) Artifact B77

Figure 5.45: Generation 1 artifacts based on B60 (Experiment 13)

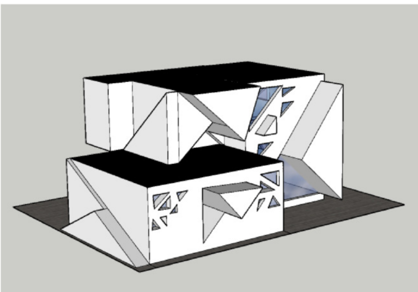


(a) Artifact B74

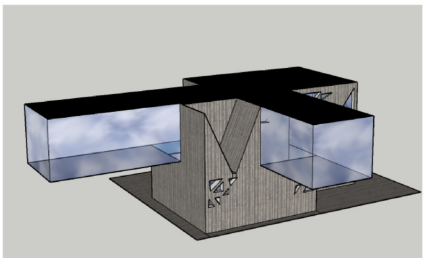


(b) Artifact B76

Figure 5.46: Generation 1 artifacts based on B64 (Experiment 13)

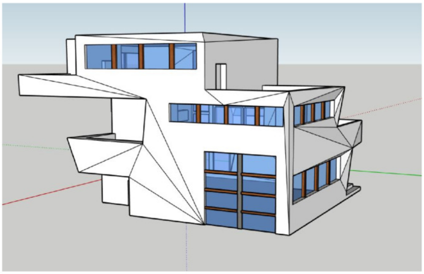


(a) Artifact B80

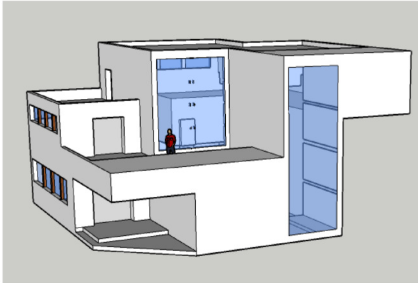


(b) Artifact B82

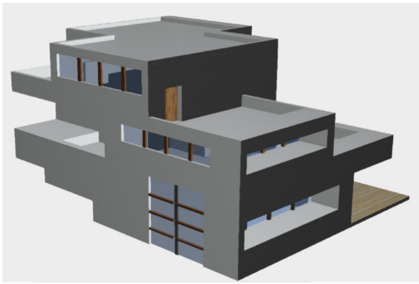
Figure 5.47: Generation 1 artifacts based on B63 (Experiment 13)



(a) Artifact B78

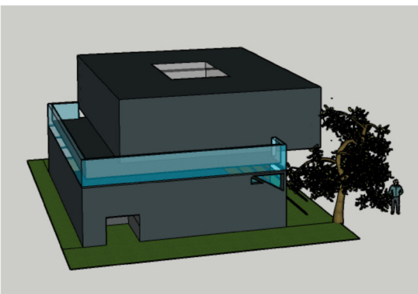


(b) Artifact B81

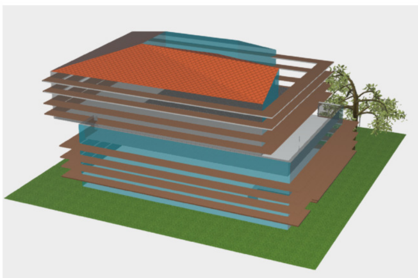


(c) Artifact B86

Figure 5.48: Generation 1 artifacts based on B66 (Experiment 13)

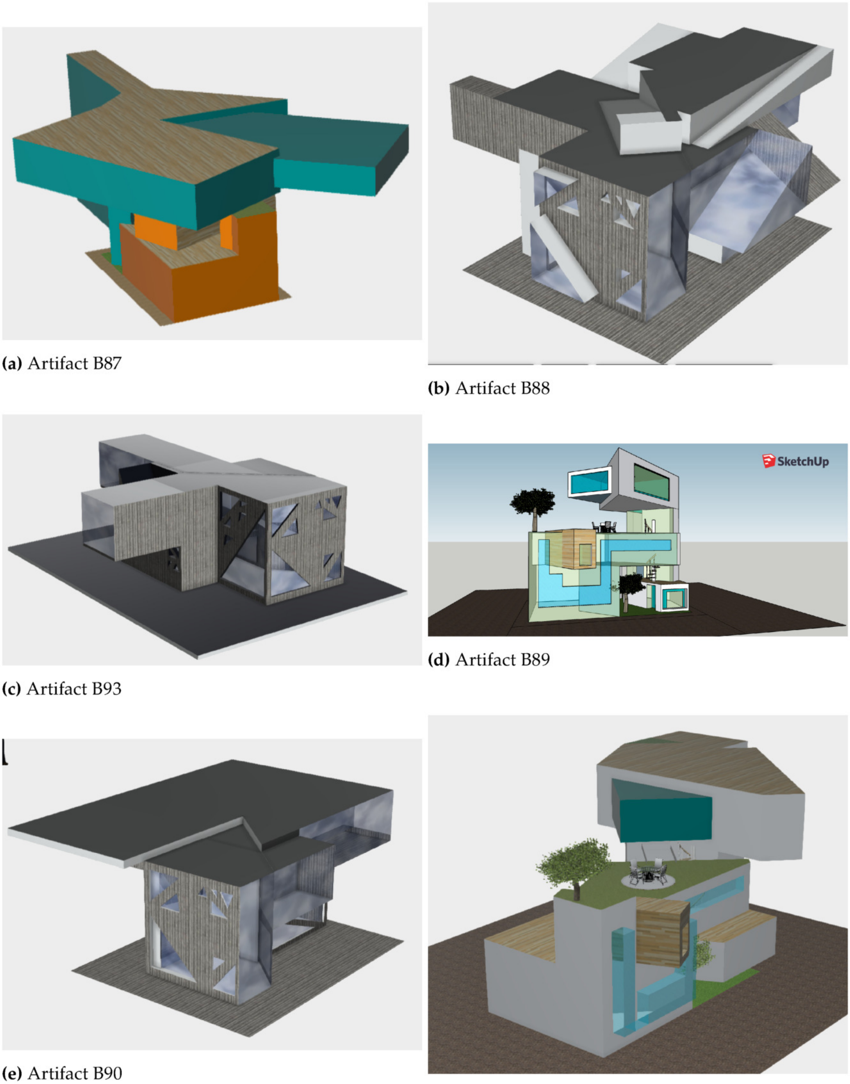


(a) Artifact B84



(b) Artifact B79

Figure 5.49: Generation 1 artifacts based on B61 (Experiment 13)



(a) Artifact B87

(b) Artifact B88

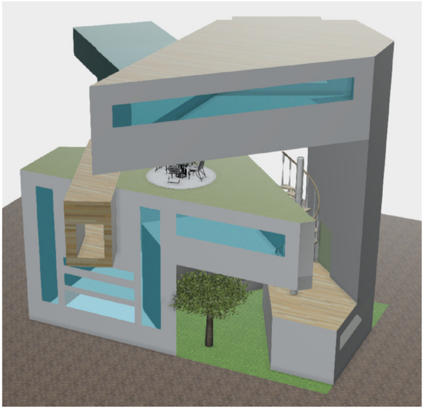
(c) Artifact B93

(d) Artifact B89

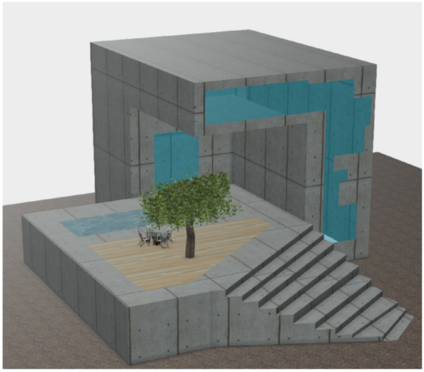
(e) Artifact B90

(f) Artifact B94

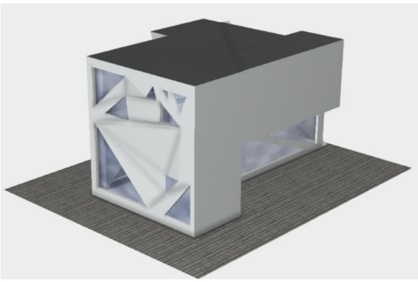
Figure 5.50: Generation 2 artifacts (Experiment 13)



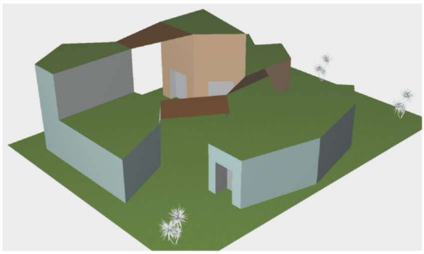
(a) Artifact B95



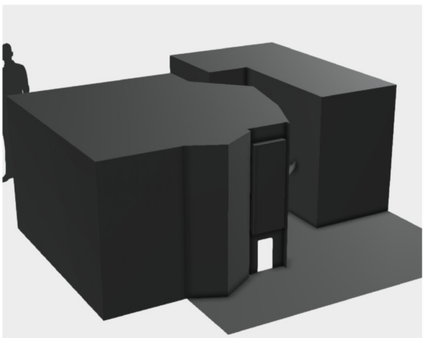
(b) Artifact B97



(c) Artifact B91



(d) Artifact B92



(e) Artifact B96

Figure 5.51: Generation 2 artifacts (Experiment 13)

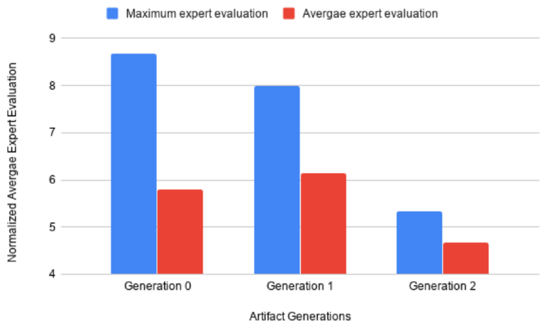


Figure 5.52: Maximum and average expert ratings for following artifact generations (Experiment 13)

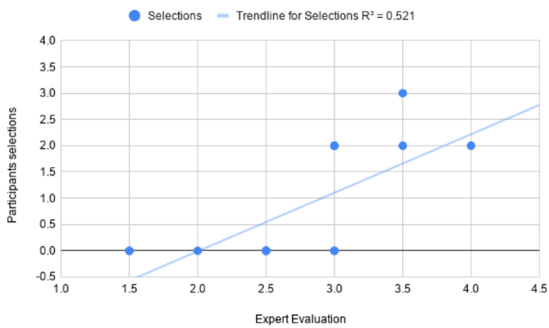


Figure 5.53: Artifact selection and expert rating per artifact (Experiment 13)

5.3 Detached House Interior Design (Project 3)

In this part of Project 3, the goal was to design the interior of the existing house (see Section 5.2). The house was selected since its design allowed for an easy division (the house was divided into four levels). In addition, we had access to its design documents.

The brief included the following required program:

1. Sitting area with sofas, coffee table, and television.
2. Dining room with a table for eight people.
3. Kitchen with a worktop, upper cabinets, wall cabinets, work island, dining table for four people, fridge, stove, and washbasin.
4. Bedrooms with a bed, closet, and desk.
5. Master bedroom with a king-size bed, desk, and closets.
6. bathroom with a bathtub, shower (if possible) sink, and a toilet.
7. Laundry room with a washing machine, dryer, sink, and closet.
8. Guest toilet with a sink and a toilet.

5.3.1 Interior Design Experiment

In Experiment 14, we explored a crowdsourcing workflow using AutoCAD, a well-known software used for technical drafting architectural artifacts.

Unlike previous experiments that were given in class, this experiment was conducted asynchronously, and

Figure 5.54: The provided building envelope renderings (Experiment 14)

participants performed it at home. We wanted to learn from the experiment about the differences in quality and quantity between the two kinds of task execution. It is essential since crowdsourcing micro-tasks are mostly performed asynchronously.

Aims

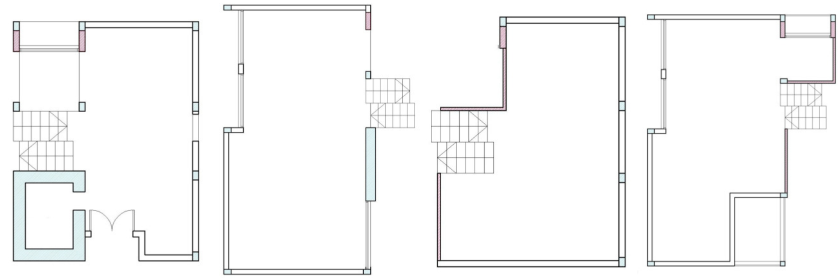
This experiment aimed to 1) explore the feasibility of an interior design task; 2) investigate a task based on spatial divisions; 3) explore the possibility to execute a task asynchronously, and 4) find out the technical challenges using AutoCAD with technical drawings.

Method

In order to prepare artifact fragments for the next experiment, a small one-family detached house was selected. The house could be simply divided, as it was made of four levels that divided the building naturally.

On December 25, 2018, we sent a task via email to 15 students. The task email included a brief, task instructions, requirements, images, and 5 AutoCAD files — one for each level and a merged file. The participants were required to design the interior of the building using these files. The participants were allowed to design every level as many times as they liked. Finally, they were requested to send the plans after two days.

The participants were provided with five AutoCAD files connected using 'external references' — a mechanism that allows for including one file inside another. One file was the 'main' plan that had references to four other sub-files. Each file included a level plan, and the main file displayed an assembly of all files. File 0 was the entry-level plan at height 0.00m, File 1 was the garden level at height 1.00m, File 2 was the bedroom level at height 2.80m, and File 3 was a bedroom level at height 3.80m (Figure 5.55). The participants were required to



(a) Level 0 **(b)** Level 1 **(c)** Level 2 **(d)** Level 3

Figure 5.55: Provided empty building envelope plans (Experiment 14)

edit only the embedded files and organize the different spaces within them.

To achieve a standard that would allow us to assemble AutoCAD files by different authors, we added the following three instructions:

1. No walls should be removed from the building envelope, but additional openings may be added.
2. Interior walls should be 10cm thick. Exterior walls need to be 20cm thick.
3. The layer that needs to be used for each kind of line: bricks, concrete, furniture, view lines, overhead lines, windows, doors, and hatches.

After receiving all the files from the participants via email, we compared them to the actual plan of the house. The plans were made over a year by an architect in collaboration with the clients. The actual design has the following distribution of spaces:

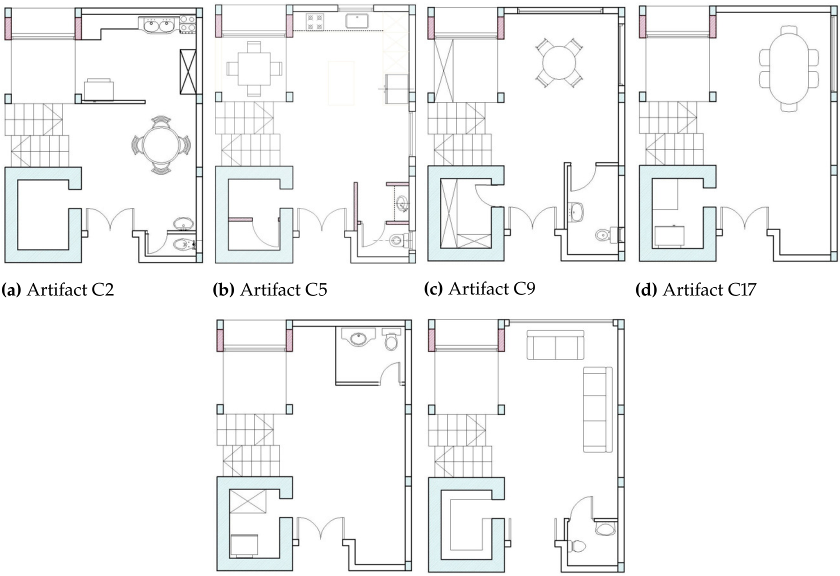
- Level 0 - Kitchen, dining area, and guest WC
- Level 1 - Sitting area and a dining room
- Level 2 - 2 Bedrooms and a bathroom
- Level 3 - Office, master bedroom, and laundry

Generated Data and Designs

Nine out of 14 students provided each between one and eight plans. A total of 30 plans were provided with the distribution outlined below.

Level 0 (6 plans)

- C2, C5 - Kitchen and dining area, and guest WC
- C9 - Dining area and guest WC
- C17 - Dining room
- C21 - Guest WC
- C29 - Sitting area and guest WC



(a) Artifact C2 **(b)** Artifact C5 **(c)** Artifact C9 **(d)** Artifact C17

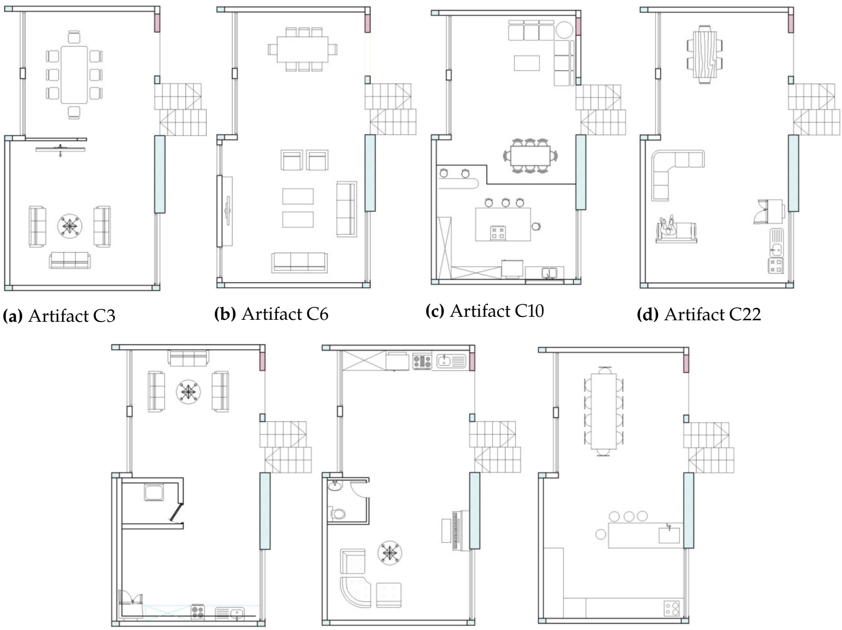
(e) Artifact C21 **(f)** Artifact C29

Figure 5.56: Level 0 plan artifacts (Experiment 14)

Level 1 (7 plans)

- C3, C6 - Sitting area and dining room
- C10, C22 - Kitchen, dining area, and sitting area

- C13 - Kitchen, pantry and sitting area
- C18 - Kitchen, guest WC and sitting area
- C30 - Kitchen and dining room



(a) Artifact C3 **(b)** Artifact C6 **(c)** Artifact C10 **(d)** Artifact C22

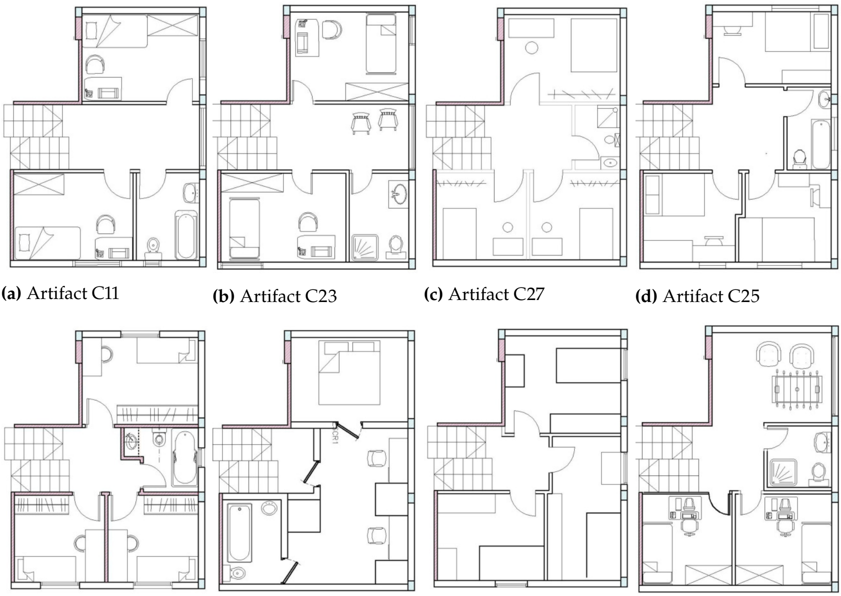
(e) Artifact C13 **(f)** Artifact C18 **(g)** Artifact C30

Figure 5.57: Level 1 plan artifacts (Experiment 14)

Level 2 (8 plans)

- C11, C23, C27 - 2 bedrooms and a bathroom with a lobby.
- C25 - 2 bedrooms and bathroom
- C7 - 3 bedrooms and bathroom
- C14 - Parents' suite
- C15 - 3 bedrooms
- C19 - 2 bedrooms, bathroom, and playroom

Level 3 (9 plans)



(a) Artifact C11 **(b)** Artifact C23 **(c)** Artifact C27 **(d)** Artifact C25

(e) Artifact C7 **(f)** Artifact C14 **(g)** Artifact C15 **(h)** Artifact C19

Figure 5.58: Level 2 plan artifacts (Experiment 14)

- C4, C20, C24, C26, C28 - Parents' suite and laundry room
- C12, C16 - Parents' suite, bedroom, laundry room, and lobby
- C5 - Parents' suite, working area, and laundry room
- C1 - Parents' suite and office with a lobby

A total of 15 survey responses were collected. Regarding the complexity of the task, only two (13%) participants thought that the task was complex (2), while four (26%) and five (33%) thought it was clear or very clear, respectively. None of the participants thought the task was too complicated. Also, two-thirds of the participants thought that the division of the house did not limit their artistic expression.

Since this task was among the final experiments in the first workshop, we wanted to know which tool the students preferred to use for the tasks. Most participants preferred

AutoCAD (57%), five participants preferred sketching (36%), and only one preferred SketchUp (7%).

Also, the survey showed that the reason why only 11 participants sent artifacts was related to having other presentations related to other courses.

Analysis of Design and Data

The task was successful in generating multiple and valid plan artifacts. However, 36% of the participants did not provide artifacts. The participants told us that the main reason for not providing the artifacts was study-related pressure.

The participants used the following two strategies: one strategy was to set the kitchen on Level 0 and the living room on Level 1, while the other strategy was to set the kitchen on Level 1 and the living room on Levels 0 or 1. The rest of the house organization was straightforward. On Level 2, the bedrooms were placed, and on Level 3, the master suite. However, we identified a significant quality difference between the different plans.

The survey results showed that the participants found the task clear and easy; however, this could be explained by the fact that most participants preferred using AutoCAD over sketching and SketchUp. This finding suggests that the potential of AutoCAD-based design tasks.

Around a third of the participants thought that the house-plans' division had a negative effect on their artistic expression. However, we acknowledge this critique only partially because in real-world interior-design is ofter limited by levels, building envelope and vertical transportation as well as structural elements. This issue has to be explored further to find a balance between artistic expression and teamwork.

The participants successfully created different design options using AutoCAD software. The generated plan fragments had notable quality differences, where some plans were of higher quality while others were of lower quality. We also identified that most repeating options were very similar to the actual plan of the house.

Conclusions

The results suggest the following conclusions:

- It is possible to generate various interior-design plans using a micro-task.
- The spatial division worked well and did not limit the production of artifacts.
- Asynchronous tasks have a lower participation rate but are successful in producing artifacts.
- Using the AutoCAD software, which helped the students to produce valid interior design plans, is a good option for drafting 2D artifacts.

5.3.2 Artifact-Set Generation Experiment

Experiment 15 evaluated a new method the participants composed an artifact-set. Upon generating plans for each level of a house in Experiment 14, they tried to merge them using a combination method in this experiment.

Aims

The aims of this experiment were to 1) evaluate the artifact-set generation task; and 2) compare set-selection, artifact-selection, and artifact-rating methods to select the best artifacts.

Method

New task screens were developed for this experiment. The set generation task required an artifact selection screen that displayed the artifacts grouped by levels (see Figure 5.60). The second screen displayed a 'rating screen' for each artifact-set that included a display of a set and an evaluation form (see Figure 5.61). The 'Sets' and 'RateSets' tables were added to the database to store the sets and set ratings.

The experiment was conducted on January 1, 2019, with 14 students.

First, the participants were shown the artifact selection screen with 30 artifacts and asked to choose one artifact from each level to create a 'set.' The students had to ensure that the plans were aligned and represented the "best possible set".

The generated sets were displayed on the 'rating screen.' For each artifact-set, the participants had to provide an evaluation by answering the following three questions:

1. Are some of the plans conflicting? Possible answers: Yes / No
2. Does this set fulfill the functional requirements? Possible answers: Yes / No
3. How do you evaluate the total design quality of the set? Possible answers: Bad, Low, Average, Good, or Excellent.

Finally, the participants answered a survey about their experience.

Each participant generated a set, so a total of 14 sets were generated. However, 12 of the sets were unique.

Artifact-set identified as 5-6-7-8 was selected three times. The remaining sets were selected only once.

The selection frequency of the artifact in the sets was counted (Figure 5.63) Further detail on the sets is provided below.

- ▶ Artifact C6 was selected most frequently (7 times).
- ▶ Artifacts C5, C7, C8, and C25 were selected five times.
- ▶ Artifact C19 was selected four times.
- ▶ Artifacts C4, C29, C10, C24, and C2 were selected three times.
- ▶ Artifacts C9, C30, C12, C13, C20, C21, C22, C3, and C28 were selected once.

Figure 5.60: Artifact-set composition screen (Experiment 15)

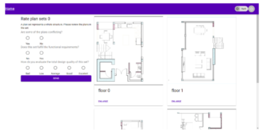


Figure 5.61: Artifact-set rating screen (Experiment 15)

The participants generated a total of 524 rating items. The average rating for each set is shown in Table 5.22.

Table 5.22: Selection distribution (Experiment 15)

Set	Conflict	Function	Quality	No. selections
2-24-6-19	71%	64%	100.00%	1
2-6-25-4	71%	71%	87.76%	1
22-21-20-19	67%	33%	58.10%	1
29-10-19-12	67%	60%	86.67%	1
29-13-25-8	57%	64%	69.39%	1
29-30-7-28	71%	100%	91.84%	1
5-3-25-24	57%	100%	89.80%	1
5-6-25-4	67%	73%	96.19%	1
5-6-7-8	60%	80%	96.19%	3
6-2-19-8	80%	80%	92.38%	1
9-10-25-24	73%	87%	84.76%	1
9-10-7-4	73%	73%	88.57%	1

We received a total of 15 survey answers. Most participants stated that the task was simple (73%), while the remaining students thought it was average. Likewise, most participants thought that the selected artifact-set (5-6-7-8) was better than the set they had assembled.

Analysis of Design and Data

The results of the analysis of the set-selection data showed that the set 5-6-7-8 was the preferred set, since it was selected most frequently, and the survey results showed a high agreement between the participants.

Furthermore, we computed the artifact-selection score by calculating the score for each possible combination based on the individual artifact selections. This resulted in 540 possible artifact-sets that scored between 7 and 77 selections. The distribution of the scores resembled a normal distribution with an average of 12.13, a median of 12, and a standard deviation of 3.01. The top-rated sets were 5-6-7-8 and 5-6-25-8 (the score of 22), and 5-6-19-8 (the score of 21).

We reviewed these artifacts (C7, C25, and C19). Their analysis showed that artifacts C7 and C25 had a similar design. However, artifact C7 was superior since the plan was more efficient and provided larger bedrooms. A review of the remaining artifacts and sets showed that the artifact-set 5-6-7-8 was superior to all other artifacts.

The most frequently selected set (5-6-7-8) was made entirely from artifacts made by one participant. Therefore, it is likely that the participants identified all parts of a specific consensual design created by a sound designer. This implies that fragmenting and merging fragments as a creative method has a limited effect.

The student who designed artifact set (5-6-7-8) admitted that she got professional help from her sister, an interior designer, since the task was performed at home. This could explain the difference in quality between these and other artifacts. However, the participants identified the relatively high quality of the design.

The analysis of the rating results showed a low correlation between the indications ‘conflict’, ‘function’, ‘quality’ and the actual set-selections ($R^2 < 0.06$) and between the average score and the calculated selection score ($R^2 = 0.19$) (see Figure 5.63) The quality rating had a low correlation with the selection score ($R^2 = 0.43$); however, ‘function’ and ‘conflict’ were negligible. Moreover, the most frequently selected set (5-6-7-8) was rated only in the fifth place.

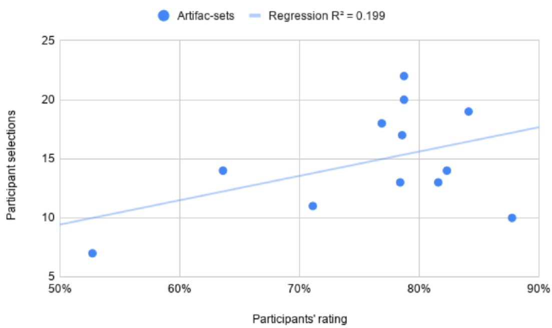


Figure 5.63: Artifact ratings vs. artifact selection score (Experiment 15)

A comparison of the results of set-selection, artifact-selection, and artifact-rating methods showed that the set-selection provided the most accurate results. Artifact selection provided some high-grade options and may serve as an effective method to select multiple artifact-sets. The artifact rating method resulted in problematic ratings that we could not accept because they did not indicate which artifact is of higher design quality.

Conclusions

Based on the results, the following conclusions were made:

- The artifact-set task was feasible and generated high-quality sets.
- In determining the quality of artifacts, rating mechanisms are less distinct than selection mechanisms. Therefore, it is advisable to use 'selection' as the primary method for identifying the quality of artifacts in the design process.
- Since a professional helped making the selected artifacts, it can be concluded that tacit design knowledge is an important factor in design tasks. Therefore, professional architects should participate in design tasks.
- The participants identified the designs made by a professional, suggesting that evaluating design can be performed by participants with limited design education.

(a) Artifact C4
(b) Artifact C20
(c) Artifact C24
(d) Artifact C26
(e) Artifact C28
(f) Artifact C12
(g) Artifact C5
(h) Artifact C16
(i) Artifact C1

Figure 5.59: Level 3 plan artifacts (Experiment 14)

5.4 Idan Refreshment Station (Project 4)

Project 4 was performed in the second semester (2019) with nine architecture students. The participants had different levels of experience: three students were in their second year, four were in their third year, one was in the fourth year, and one was in the fifth year.

We designed a new project to support the experiment plan. The project had to be in a distant location, not accessible to the participants geographically. We chose to design a small and simple building that would include basic features like water supply, sanitation, and electricity. The project goal was to design a desert tourism center in 'Idan,' a small locality in the Arava desert in Israel.

The brief was as follows: "A new desert tourism center needs to be planned. It will be used by visitors and residents of the area. The building will be located near the village 'Idan,' in the northern Arava desert. The building should have a store that will sell drinks, food, various products for travelers and provide travelers with information on the routes and businesses in the area. The building will be located at the village gate."

Project Objectives

- A place to refresh before and after trips
- A meeting place for the local community
- Source of tourist information

Project Stakeholders

- Tourists and travelers
- Local community
- Employees

Site location

Address: Lat: 30.804178, Lon: 35.295642

Figure 5.64: Satellite image of Idan in the northern Arava desert (source: Google maps). The project location is marked in red.

With an updated version of the software, we specified the brief with the following qualitative and quantitative requirements:

Business Requirements

- Desert narrative design: required
- Seats: minimum 20
- Storage: between 5-8 m^2
- Kitchen: minimum 8 m^2
- Self service area: minimum 20 m^2
- Restrooms: required
- Parking lot: required
- Sales counter: required

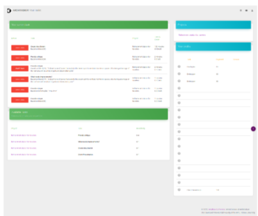


Figure 5.66: Main dashboard where tasks can be viewed

User Requirements

- Accessible facility: required
- Cover from rain and winds: required
- Shading from the sun and ventilation: required

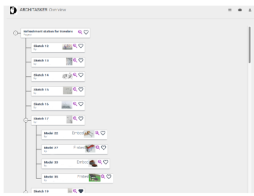


Figure 5.67: Design process tree screen

Technical Requirements

- Made of materials that last at least 30 years: required
- Total area: between 40-60 m^2
- Height: 3-5 m

Figure 5.65: Satellite image of the project site in Idan (source: Survey of Israel). The project location is marked in red.

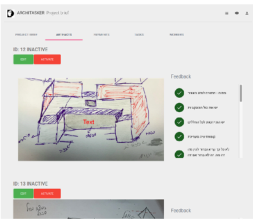


Figure 5.68: Artifacts administration screen

During the semester break, most of the components of the previous software were replaced. The changes in the software considered the conclusions from the previous experiments and provided a technological infrastructure for conducting experiments by the plan. All application screens were updated.

The main features of these changes were as follows:

- The possibility of having several projects in parallel, allowing for multiple concurrent experiments and archiving experiment data, was added. This allowed for a control group (see Figure 5.66).
- Users could now be associated with projects and roles: designer, reviewer, set-selector, etc. This was useful for the automatic assigning of tasks to groups of participants.
- Automatic task generation and assignments according to different conditions were added. A strategy programming pattern allowed for selecting algorithms in run-time. This allowed us to explore

different task assignment algorithms, require participants to resolve tasks, and better control the workflow.

- Administration interface with task management, participants' management, and artifact management was added. This helped us to better manage the experiment (see Figures 5.67, 5.68, 5.69, and 5.70).

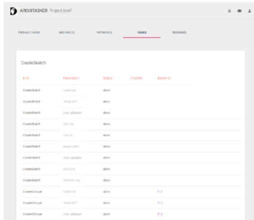


Figure 5.69: Task administration screen

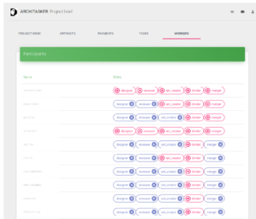
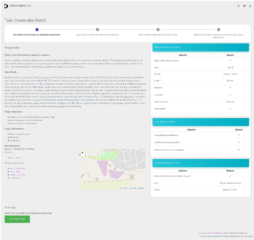


Figure 5.70: Participant administration screen (Names are blurred for privacy)

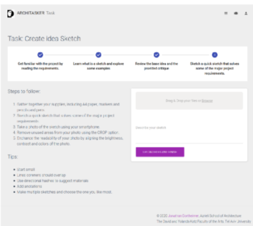
The database implementation included the following tables (see database diagram in Figure A.24):

- Accounts: Participant information including name, email, password, experience years, and birth year
- Artifacts: Artifact information, including description and artifact kind. Every artifact is related to a project, account, and task.
- ArtifactCritiques: Text-based review data that is related to an artifact, account, and task
- ArtifactUploads: Metadata of uploaded files. Evert upload is related to an artifact and an account
- Artifactsancestors: Hierarchical relations between artifacts that represent the design process as a tree
- Projects: Project brief information. Every project has an account as an owner.
- ProjectLinks: Optional links to web resources relevant to the project brief
- ProjectRequirements: Project requirements for each project, including the description, kind, and acceptance range
- ProjectRoles: The relationships of projects account roles (e.g., designer, reviewer, set creator, merger, etc.)
- RequirementCritiques: Participants' artifact and rating project requirements
- SetArtifacts: Participants' selection artifact sets
- Tasks: Assigned tasks that include task kind (sketch, set, review, etc.), status (to do, in progress, done), payment, and time estimate. Tasks are related to a project, account, base artifact, and generated artifact.
- Sessions: Current user sessions for authentication



(a) Task brief screen

(b) Task example output screen



(c) Task screen

Figure 5.71: Task screens (Experiment 16)

purposes.

5.4.1 Conceptual Sketch Task Experiment

Experiment 16 explored the new and improved conceptual sketch task.

Aims

The experiment aimed to evaluate the improved process for creating conceptual sketches from the brief.

Method

Compared to the previous sketching tasks, the artifact generation task was improved to include the following three new screens: Brief, output example, and task steps page.

The brief screen contained a project description, client description, maps, stakeholders, and location descriptions (see Figure 5.71a). The brief also included quantitative and qualitative requirements and links to weather information, Wikipedia, and the official website of Moshav Idan.

The output description screen contained a variety of sample sketches, a video demonstrating an architect creating sketches, and a verbal explanation (see Figure 5.71b). The task screen included a description of the steps for making a sketch, capturing it with a smartphone, editing the photo, and uploading it to the software (see Figure 5.71c).

The experiment was conducted on March 26, 2019, with seven students. The participants were asked to create a sketch of a building that would meet the requirements of the brief sent to participants 24 hours before the experiment.

The new system provided the participants with a page with multiple sketch examples and a video explaining what a sketch is and how to perform the task. The participants drafted sketches, photographed them with their smartphones, edited the images, and uploaded them to the software. Then a survey was conducted to learn about the participants' experience. Finally, the quality of the artifacts was evaluated by experts.

Generated Designs and Data

All nine participants succeeded in generating sketches (see Figure 5.72 and 5.73). The list of artifacts, participants' experience years, and expert evaluation is shown in Table 5.23. Four participants answered a survey about their experience while performing the task.

Artifact	Experience (Yrs)	Expert evaluation
D12	2	1
D13	2	1
D14	3	1
D15	2	2
D16	3	1
D17	3	1
D19	4	3
D20	5	5
D21	3	2

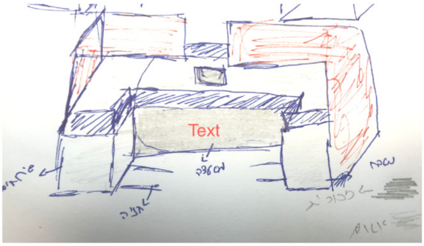
Table 5.23: Generated artifacts, participant experience, and expert evaluation (R^2 = 0.68) (Experiment 16)

Analysis of Design and Data

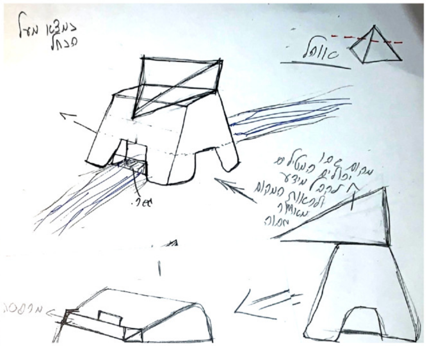
There was a strong relationship between the artifact design quality evaluation and the participants' experience (R^2 = 0.68). The artifact with the highest design quality, D20, was created by the most experienced participant (5 years), followed by D19, which was made by a slightly less experienced participant (4 years). The artifacts (D20, D19, and D15) expressed significant architectural potential.

We analyzed and classified the artifacts into the following five groups:

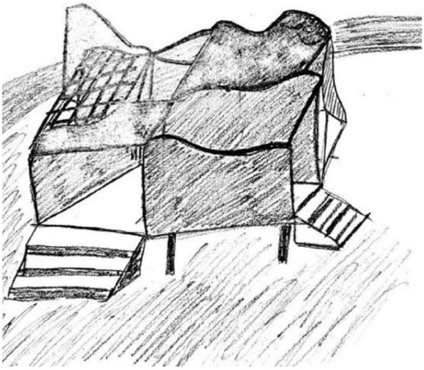
- Courtyard buildings (Artifacts D12 and D16). Various buildings surround a courtyard.
- Tent buildings (Artifacts D13 and D20). Artifact D13 was a two-story building inspired by desert tents allowing views from the top level. Artifact



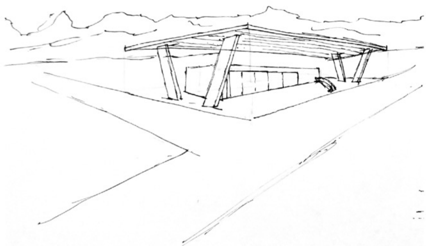
(a) Artifact D12



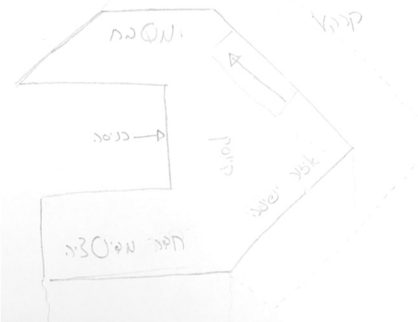
(b) Artifact D13



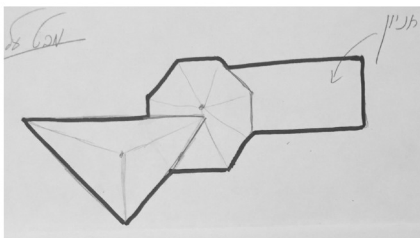
(c) Artifact D14



(d) Artifact D15



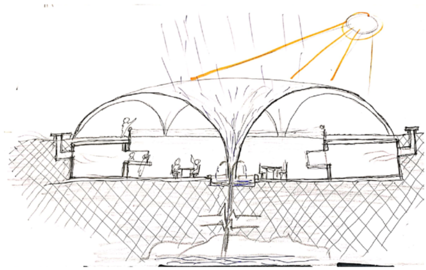
(e) Artifact D16



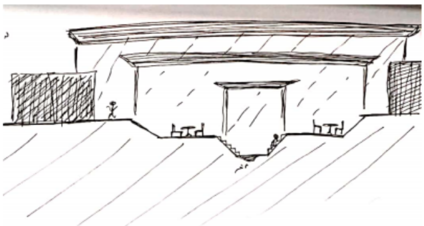
(f) Artifact D17

Figure 5.72: Artifacts (Experiment 16)

(a) Artifact D19



(b) Artifact D20



(c) Artifact D21

Figure 5.73: Artifacts (Experiment 16)

D20 was a round underground structure with a circular tent tensile structure.

- Wave buildings (Artifact D14 and D19). Artifact D19 was a glass facade pavilion covered by waveform roofs. The pavilion was divided into three spaces. Artifact D14 was a one-story building based on small pillars. The building's inner and outer walls were rounded and suggested a wavy design.
- Big roof building. Artifact D15 was a simple rectangle building shaded by a separated large roof.
- Polygonal buildings (Artifacts D17 and D21). Artifact D17 was a building made out of rectangle, hexagon, and triangle prisms covered by a gabled roof.

The participants reported having enjoyed creating sketches and having had "many ideas." The mean score of the question 'How clear was the sketch task?' was 4.75 (out of 5), meaning that participants thought the task was clear.

The mean score of the question 'How complex was the sketch task?' was 3.25 (out of 5). Despite simplifying requirements and sending the brief 24 hours before the experiment, the participants suggested adding more time and providing a surveyor map to facilitate the design task.

Conclusions

Based on the results, we concluded that the new task format was clear. All participants succeeded in generating sketches in line with their levels of experience, so the task did not limit their skills. The results of the survey showed significant improvements in participant satisfaction.

5.4.2 Review Task Experiment

In Experiment 17, a new and improved review task was evaluated.

Aims

The aims of the experiment were to 1) perform a new 2-screen review task; 2) evaluate the new requirement rating part; and 3) measure the valid review rate of the reviews.

Figure 5.74: Review task screen (Experiment 17)

Method

A new 2-screen review task user interface was developed. The interface included a brief and a review screen. The review screen displayed the reviewed artifact files and a form made up of the following three parts. In the first part, a question was displayed for each project requirement, and the participants had to choose if the presented artifact fulfilled the requirement. In the second part, the participants had to answer the following question "What do you like about this design?". Finally, in the third part, the participants had to answer the question "How would you improve this design?" (see Figure 5.74)

The experiment was conducted after Experiment 16 on March 26, 2019, with the participation of eight students. Each participant was assigned eight review tasks, one for each artifact, and fulfilling all tasks was mandatory.

Upon completion of the task, the participants responded to a survey about their experience.

Generated Designs and Data

The students completed a total of 64 review tasks. Each participant generated 16 review items. We analyzed the content of the reviews and removed the reviews that did not include meaningful text. The remaining reviews were counted, and the validity rate was calculated based on the valid and the potential review count. A total of 119 out of 128 review items (i.e., 92.96%) were valid.

The task completion duration ranged from 12:18 to 39:54 min (mean = 22:51 min). Therefore, on average, each review item took 2:30 min to complete.

A total of 1008 requirements ratings were generated. The average scores for each artifact and requirement are shown in Table 5.24.

Table 5.24: Requirements average rating (Experiment 17)

Requirement	D12	D13	D14	D15	D16	D17	D19	D20	D21	Requirement Average
Desert narrative design	0.38	0.88	0.50	0.50	0.00	0.50	0.33	1.00	0.71	0.53
Seats	0.63	0.63	0.63	0.75	0.88	0.86	1.00	1.00	0.86	0.80
Storage	0.75	0.25	0.75	0.38	1.00	0.50	1.00	0.50	0.86	0.66
Kitchen	1.00	0.50	0.88	0.75	1.00	0.71	1.00	0.50	1.00	0.82
WC	1.00	0.38	0.75	0.63	0.88	0.88	1.00	0.38	0.75	0.74
Parking lot	1.00	0.13	0.13	0.63	0.13	0.50	0.25	0.88	0.63	0.47
Self-service area	0.13	0.43	0.50	0.38	0.38	0.75	0.71	0.63	0.63	0.50
Sale counter	0.50	0.25	0.50	0.63	0.88	0.57	0.75	0.75	0.50	0.59
Disabled-accessible	0.38	0.25	0.25	0.75	0.75	0.43	0.63	0.13	0.00	0.39
Cover from rain and winds	0.75	0.88	0.88	0.88	0.75	1.00	1.00	0.71	0.63	0.83
Protection from the sun	0.63	1.00	1.00	1.00	0.75	1.00	0.63	0.88	0.75	0.85
Sustainable materials	0.63	0.88	0.88	0.75	0.50	0.75	0.75	0.50	1.00	0.74
Total area	1.00	1.00	0.88	0.88	1.00	1.00	1.00	1.00	1.00	0.97
Height	0.88	1.00	0.88	1.00	1.00	1.00	1.00	1.00	0.88	0.96
Artifact Average	0.69	0.60	0.67	0.71	0.71	0.75	0.79	0.70	0.73	0.70

Analysis of Design and Data

The task generated many valid reviews and improvements ideas. The validity rate was 92.96%, which is slightly higher than in previous experiments (see Table 5.25). This may result from changing the review questions in experiment 9 from three questions to two.

Table 5.25: Review sessions validity rate comparison (Experiment 17)

Review Experiment	No. Questions	Artifacts to review	Valid Items	Items Count	Validity Rate
Experiment 3	3	12	69	93	74.19%
Experiment 6	3	13	196	259	75.68%
Experiment 9	3	13	277	307	90.23%
Current experiment	2	9	119	128	92.97%

The analysis of the requirement rating distribution (Table 5.24) showed that the average artifact rating score ranged between 0.60 and 0.79 (mean = 0.70). There was a minor difference between the artifact requirement rating.

Some requirements — such as the existence of the WC, seats, storage, sales counter, parking lot — were not relevant to the conceptual stage. Other requirements important for the concept stage, such as the size of the structure, as well as protection from rain and sun, were mostly applied by the designers and received high ratings with a low standard deviation. On the other hand, the requirement for a desert narrative, which was essential, was agreed only for artifact D20 and was rated high for D13 and D21. The remaining artifacts were rated low on this requirement.

The results of the survey showed that some students considered providing multiple reviews for all artifacts to be exhausting and repetitive. Therefore, this means that it would be better to provide reviews only to the selected artifacts.

Conclusions

The following conclusions were made:

1. The task was successful in generating relevant review data.
2. The review task should be executed after the selection task because otherwise, it is too exhausting and time-consuming; in addition, it is also ineffective to review artifacts that are not selected.
3. Some technical requirements are not necessary for the concept sketch stage. However, conceptual requirements and less detailed requirements such as height, size, shading, and protection from rain are essential, and their rating should be taken into consideration.

5.4.3 Selection Task Experiment

In Experiment 18, we evaluated a new selection task that replaced the rating task from previous experiments. The idea was to allow the participants to select the artifact they liked the most (instead of providing individual ratings).

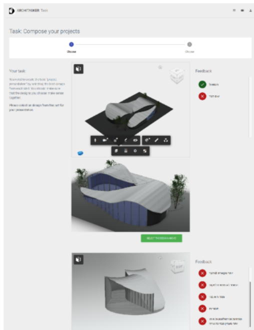


Figure 5.75: Selection task screen (Experiment 18)

Aims

The aims of the experiment were to 1) evaluate the selection task user-interface; 2) assess the task outcomes; and 3) suggest an artifact rating method that considers potential bias.

Method

Based on the conclusions of the rating and artifact-set experiment, a new selection task was developed. After the participants browsed the project brief, the selection screen was displayed (Figure 5.75).

The experiment was conducted on March 26, 2019, with nine students. Upon completion of the experiment, the participants received a new task ("Select artifacts"). All nine artifacts generated in Experiment 16 were displayed, including the reviews from Experiment 17. An average of requirement score was displayed with the reviews. The artifacts were ordered by their average fitness scores generated in previous experiments.

Generated Designs and Data

The participants finished the task and provided nine selections (see Table 5.26). Completing the task took, on average, 1:46 min.

Table 5.26: Selection distribution (Experiment 18)

Artifact	Non Author	Author	Selections	Selection Score	Requirement Score	Expert Evaluation
D12	0	0	0	0	0.69	1
D13	0	0	0	0	0.60	1
D14	0	1	1	0.5	0.67	1
D15	0	1	1	0.5	0.71	2
D16	1	0	1	1	0.71	1
D17	1	1	2	1.5	0.75	1
D19	2	0	2	2	0.79	3
D20	1	1	2	1.5	0.70	5
D21	0	0	0	0	0.72	2

Analysis of Design and Data

The results showed a correlation between the requirement scores and the number of selections ($R^2 = 0.39$) (see Table 5.26). According to expert evaluation, artifacts D19 and D20 were the best, and the selection process succeeded in identifying them. However, the participants also selected artifact D17, which was previously evaluated as low quality.

Since four out of nine designs selected their authored artifact, there was a bias in the data. This possibility was also confirmed by the ratings of artifacts D14 and D15, where the selection of the designers was the only selection. However, artifact D19 was selected the most, but not by the design author.

Since we reasoned that the designers should participate in the selection micro-task, the bias needed to be considered, particularly in small groups. To address this issue, we computed a selection score made out of the number of participants' selections. However, if a designer selected his/her authored artifact, this choice was weighted as half selection. The score was straightforward and helped to reduce the weight of biased selection. In our case, the score brought a sharper distinction between the artifacts and highlighted artifact D19 as the most favorable.

However, the selection score did not indicate the highest evaluated artifact. Based on the idea that open-source development has multiple open ends developed in parallel [2], as well as considering the results of Experiment 17, we chose multiple artifacts for further development. A threshold of 50% was set so that artifacts D17, D19, and D20 were chosen for the next step.

[2]: Carpo (2011), *The Alphabet and the Algorithm*

The results of the survey showed that the participants considered the task to be easy and intuitive.

Conclusions

Based on the results, the following conclusions were made:

1. The selection micro-task was simple and did not require much time to be performed. The workflow was clear, and the task was completed within a short time.
2. The selection micro-task indicated the higher-quality artifacts but did not identify the best artifact. Therefore, the output of the task should allow more than one artifact.
3. There was still a bias when designers also participated in the selection task. The bias can be limited using the selection score.

5.4.4 Model Generation Task Experiment with Crowd-Workers

Experiment 19 was performed with the participation of crowd-worker architects recruited using an online global work-marketplace platform (Upwork). This is the

Figure 5.76: Task example screen (Experiment 19)

first time that we experimented with professional crowd-workers.

Aims

The aims of this experiment were to 1) test whether remote crowd-workers could effectively use the software; 2) compare the output of professional crowd-workers vs. students; and 3) evaluate clarity and quality of the outputs in a task where sketches are transformed to 3D models.

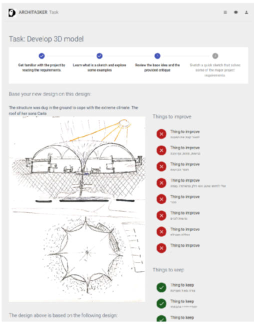


Figure 5.77: Base-artifacts screen (Experiment 19)

Method

For this experiment, we developed a new model generation micro-task that included the following screens: Brief, Output example (Figure 5.76), Base-artifact (Figure 5.77), and Task-steps (Figure 5.78).

On April 14, 2019, we posted a job on Upwork, a website for hiring freelance workers. The job posting included a short description of the task, work time limit, tolerance, a sample concept sketch, and an example of a 3D model.

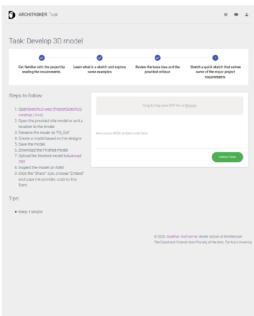


Figure 5.78: Task steps screen (Experiment 19)

The job received 38 proposals with the bid range from 7$ to 35$ per hour (mean: 16.11$ per hour). We chose the proposals with lower hourly rates (5$ to 7$). The recruited architects were from Armenia, Pakistan, Serbia, Georgia, and the Philippines. They were provided with personal user accounts and access to our software. For each job on Upwork, we assigned each participant a personal task that included all the screens described above.

The participants also answered a survey about their experience. The survey asked whether the task was clear, complicated and if the participants liked it.

Finally, expert architects reviewed the artifacts and provided a design quality evaluation.

Generated Designs and Data

A total of 5 participants generated 12 artifacts and provided five survey answers. The distribution of tasks was random. Participants 11, 12, and 13 received three tasks, participant 14 received two tasks, and participant 15, who was the last to accept the job on Upwork, received one task. It took several days for the participants to complete the micro-tasks, as some were not immediately available or were located in different time zones (the artifacts were submitted from March 28 to April 1).

Artifact	Account	Experience (Yrs)	Expert Evaluation
D22	11	4	1.43
D23	11	3	0.71
D24	14	3	2.50
D27	12	2	1.07
D28	12	2	1.43
D29	12	2	0.36
D30	14	3	5.00
D31	11	4	1.43
D32	13	4	0.71
D33	13	4	0.71
D34	13	4	0.71
D35	15	3	1.43

Table 5.27: Artifacts, participant experience, and expert evaluation (Experiment 19)

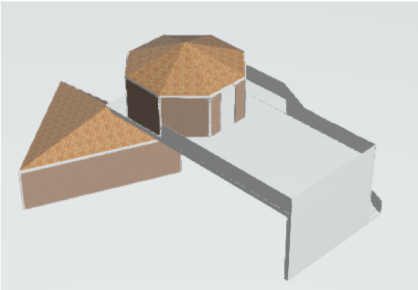
The results of the survey showed that the participants perceived the task to be clear, with an average rating on the simplicity of 4.2 out of 5. They also indicated that the task was not too simple but also not complicated (rating 3.8 out of 5). Finally, the participants rated that they liked using the software, rating it 4.4 out of 5.

In the survey, we also asked the participants to provide improvement suggestions. One participant said that working with A360[2] was not clear. Another participant suggested adding textual descriptions to the sketches so that to make them more evident.

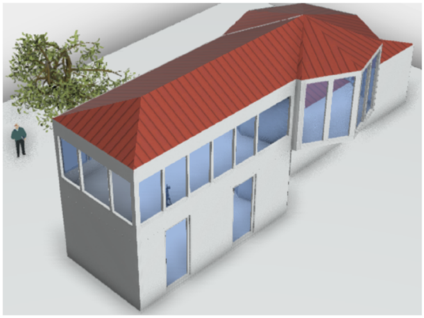
2: Autodesk's CAD storage service

Analysis of Design and Data

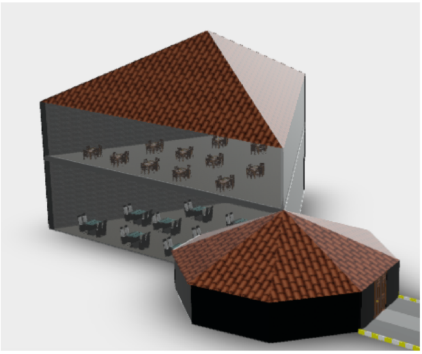
The results of the analysis of the generated artifacts can be summarized as follows.



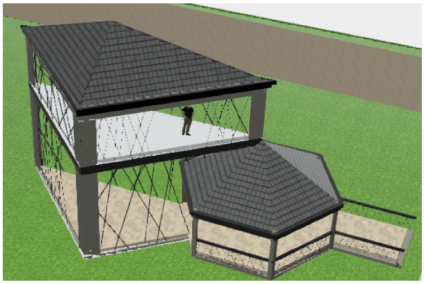
(a) Artifact D22



(b) Artifact D27



(c) Artifact D33



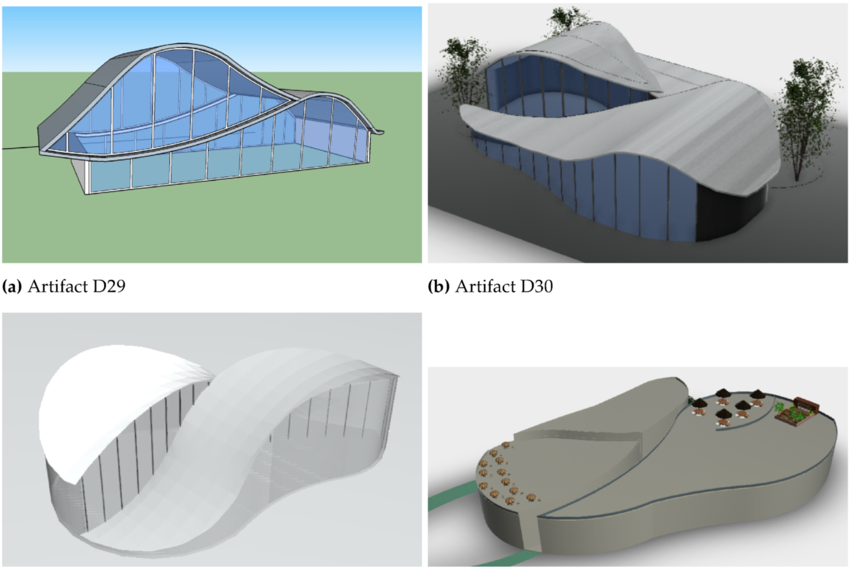
(d) Artifact D35

Figure 5.79: Artifacts based on Artifact D17 (Experiment 19)

Artifacts based on D17

1. D22 was a structure made out of hexagonal and triangular prisms, covered by a gabled roof. We evaluated the quality of this model as low.
2. D27 was a structure made of rectangular and hexagonal connected prisms, both covered by a red-tiled roof. The quality of this model was evaluated as medium.
3. D33 was a triangular and octangular structure with sloping red roofs. We evaluated the quality of this model as high.
4. D35 contained triangular and hexagonal extruded prisms covered by sloping dark roofs. The facades were glazed with irregular diagonal structural pro-

files. The quality of this model was evaluated as high.



(a) Artifact D29

(b) Artifact D30

(c) Artifact D31

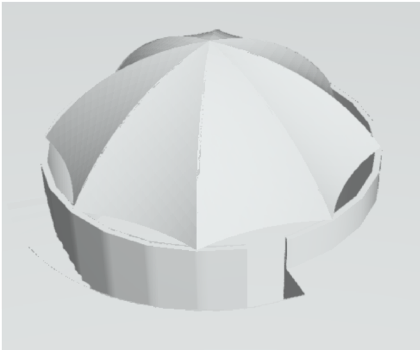
(d) Artifact D34

Figure 5.80: Artifacts based on Artifact D19 (Experiment 19)

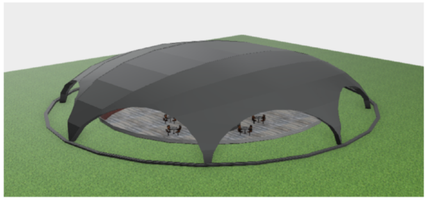
Artifacts based on D19

1. D30 was a three-drop-like structure covered by a wave-like roof with glazed facades. The quality of this model was evaluated as high.
2. D31 was a two-drop like joined structure with wave roofs and glazed facades. We assessed the quality of this model as high.
3. D34 was an extruded plan in the form of three drops. On the roof, the designer placed some furniture (see the plan sketch). The quality of this model was evaluated as low.

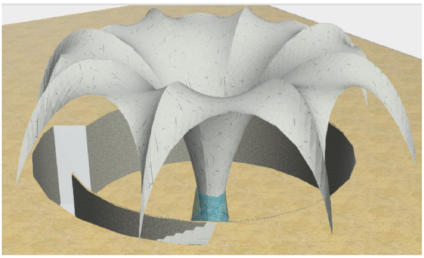
Artifacts based on D20



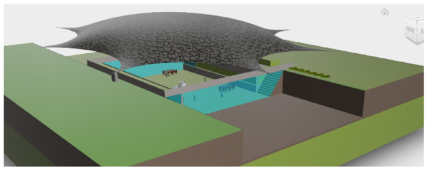
(a) Artifact D23



(b) Artifact D24



(c) Artifact D28



(d) Artifact D32

Figure 5.81: Artifacts based on Artifact D20 (Experiment 19)

1. D23 was a round structure covered by a tent-like sail dome. It was buried in the ground, and the model lacked materials. We evaluated the quality of this model as medium.
2. D24 was a buried round space with a large dome shading structure above. The dome was supported by evenly distributed eight beams. The quality of this model was evaluated as high.
3. D28 was an extreme donut sail-dome additionally supported in the center of the structure. The dome covered an excavated space. We evaluated the quality of this model as high.
4. D32 was an excavated space covered by a tent-like tension roof. The quality of this model was evaluated as medium.

We planned that each task would take an hour and expected that each task would cost between 5$ to 7$; however, due to a misunderstanding with two participants, higher rates had to be paid. The participants had between 2 and 4 years of professional architectural experience, and most had a Bachelor's degree. The education and payment information are reported in Table 5.28.

Table 5.28: Participants' education, experience, performance, and cost (Experiment 19)

Account	Location	Education	Experience (Yrs)	Hourly Rate	No. of Tasks	Actual Task Cost	Average Expert Evaluation
11	Armenia	M.Arch	4	7.00 $	3	7.00 $	1.19
12	Georgia	Student	2	6.25 $	3	6.25 $	0.95
13	Pakistan	B.Arch	4	5.00 $	3	12.50 $	0.71
14	Serbia	B.Arch	3	7.00 $	2	7.00 $	3.75
15	Philippines	B.Arch	3	5.00 $	1	25.84 $	1.43

The architects created models of variable quality. There was no relation between the quality, level of education, and actual cost; furthermore, we observed a weak correlation between the hourly rate and quality ($R^2 = 0.28$).

However, assigning multiple tasks to workers did impact the quality, and they provided lower-quality models on average. This might be because the workers did not allocate their time evenly on the tasks and invested less the longer they worked.

Conclusions

Based on the results, hiring professional architects and providing them with tasks through our software was successful. In addition, the following conclusions were made:

1. The model generation micro-task was successful in generating high-quality models; therefore, the software was effective.
2. The participants found the task to be clear, simple, and fun to perform.
3. As in experiments with students, the quality of artifacts produced by freelance architects varied.

Along with high-quality artifacts, some artifacts had a low-quality design.
4. To generate higher-quality models, tasks should be assigned individually to workers.

For the software, the following conclusions were drawn:

1. Working with external platforms is more complex and requires clearer instructions. It would be better to avoid external systems.
2. It would be easier to provide workers with an invitation link to register and assign them a task automatically.

5.4.5 Plan from Models Generation Experiment with Crowd-Workers

Experiment 21 explored the micro-task that generated 'plan outlines' from a 3D artifact. This task failed in the previous experiment; in this experiment, we tried the micro-task with crowd-workers.

Aims

This experiment aimed to 1) evaluate a simpler micro-task that generates AutoCAD plans from SketchUp 3D without dividing the plans; 2) compare the crowd-workers performance with that of the students; 3) establish whether 2D plan generation from 3D models can be performed better by crowd-workers.

Method

The experiment was conducted on April 15-16, 2019, with five crowd-workers hired through Upwork. We published a task for entry-level architects, with the following description: 'Create Autocad plans out of SketchUp model.' The job received 17 proposals with hourly rates ranging from 5$ to 80$. Two workers who participated in Experiment 19 and 3 new workers with an hourly rate between

5$ to 8.75$ were hired. The workers were from Georgia, Serbia, Egypt, Indonesia, and Chile.

The participants received credentials for an account on our software. The task was almost identical to the one in Experiment 20 (except for omitting the plan-division step so that to make the task more straightforward).

Upon completion of the task, the participants responded to a survey about their experience.

Generated Designs and Data

The participants successfully generated ten artifacts (see Table 5.29), and five survey responses were collected.

Artifact	Account	Experience	Expert Evaluation
D37	16	6	2
D38	14	3	4
D39	14	3	1
D40	16	6	2
D41	18	8	5
D42	17	12	3
D43	18	8	4
D44	14	3	3
D45	17	12	2
D46	12	2	4

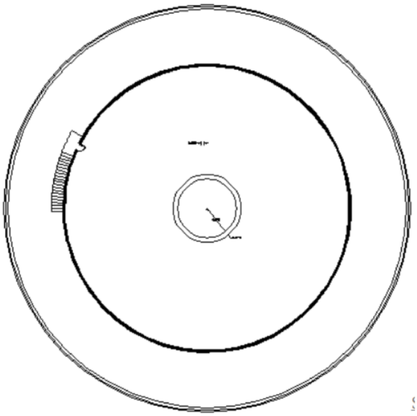
Table 5.29: Generated artifacts, experience years, and expert evaluation (Experiment 21)

Analysis of Design and Data

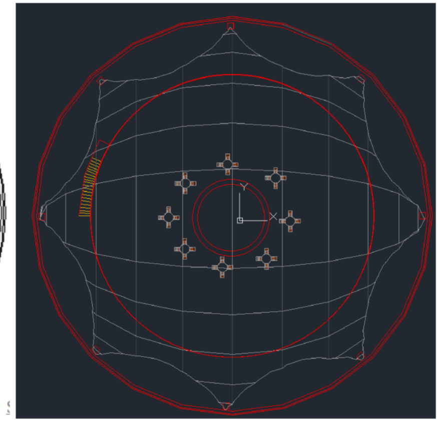
The results of the analysis of the generated artifacts are presented below.

Artifacts based on Artifact D24 (Figure 5.82)

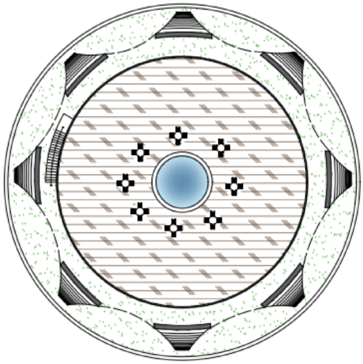
- D37: A simple section through the model
- D38: A plan that includes the projecting of the roof and furniture
- D41: A plan with textures, furniture, and roof columns with a projection of the roof as a hidden line. Also, a roof plan was provided that suggested a structural configuration.



(a) Artifact D37



(b) Artifact D38

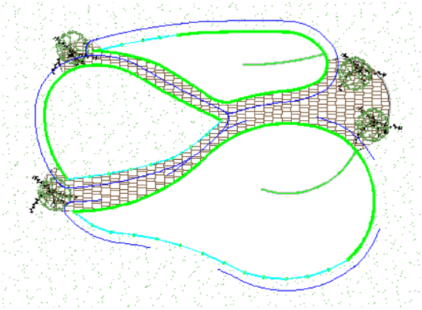


(c) Artifact D41

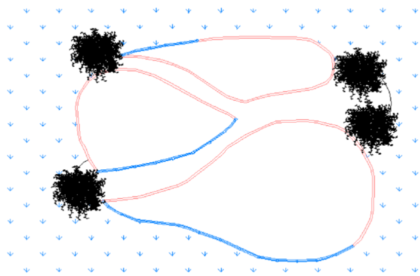
Figure 5.82: Artifacts based on Artifact D30 (Experiment 21)

Artifacts based on Artifact D30 (Figure 5.83)

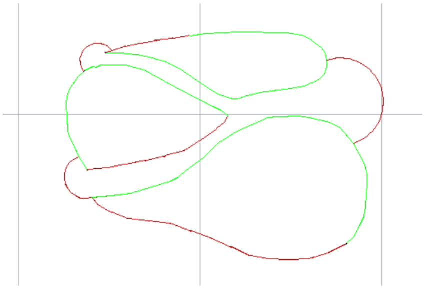
- D43: A plan with textures and trees, roof outline as a hidden line, and a roof plan.
- D44: A plan with textures
- D45: A simple plan without additions



(a) Artifact D43



(b) Artifact D44



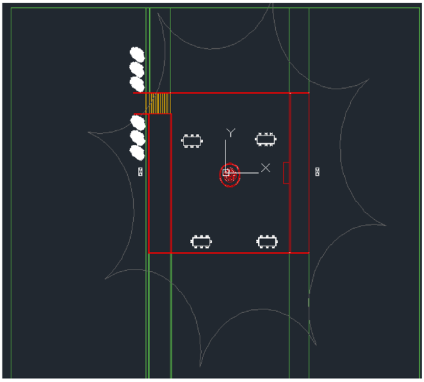
(c) Artifact D45

Figure 5.83: Artifacts based on Artifact D30 (Experiment 21)

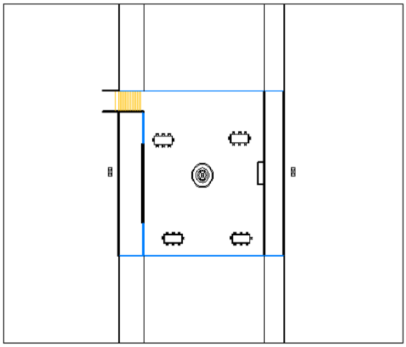
Artifacts based on Artifact D32 (Figure 5.84)

- D39: A plan with furniture with a projection of the roof
- D40: A simple section with furniture
- D42: A simple section with furniture
- D46: A plan with furniture and fauna

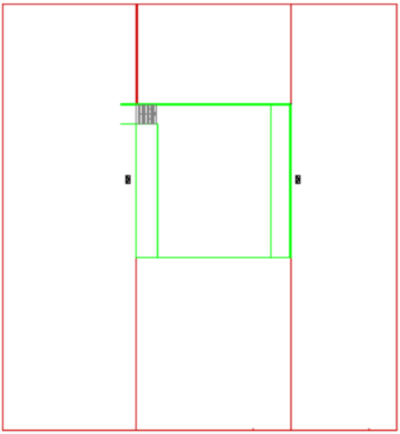
According to the results of the survey, most of the participants thought the payment was fair. We also analyzed the participants' performance based on expert evaluation, payment, and experience (see Table 5.30). As in previous experiments, there was no correlation between the participants' experience and expert evaluation. However, strong correlations between the hourly rate and quality (R^2 = 0.79) and between the actual task cost and expert evaluation (R^2 = 0.66) were observed.



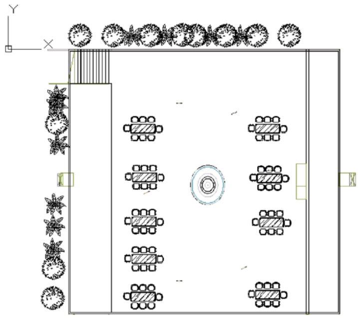
(a) Artifact D39



(b) Artifact D40



(c) Artifact D42



(d) Artifact D46

Figure 5.84: Artifacts based on Artifact D32 (Experiment 21)

Table 5.30: Participants' education, experience, performance, and cost (Experiment 21)

Account	Location	Education	Experience (Yrs)	Hourly Rate	No. of Tasks	Actual Task Cost	Average Expert Evaluation
12	Georgia	Student	2	8.25 $	1	8.75 $	4
14	Serbia	B.Arch	3	7.00 $	3	5.00 $	2.67
16	Egypt	B.Arch	6	5.00 $	2	7.08 $	2
17	Indonesia	B.Arch	12	5.00 $	2	3.50 $	2.5
18	Chile	B.Arch	8	8.00 $	2	17.33 $	4.5

The survey showed that the participants found the use of our software clear. One participant reported an issue with converting the units when exporting the SketchUp models to AutoCad. Another participant suggested using a better vocabulary to describe the task requirements.

Conclusions

All freelance architects successfully completed the task within the given timeframe. The following conclusions were made:

1. The micro-task was successful in generating high-quality plan artifacts from the provided 3D models.
2. This micro-task required specific expertise using the CAD software. Therefore, for this task, it is important to select expert participants.
3. We received better-quality output from more expensive workers who invested more time into the task.

5.4.6 Plan Division, Improvement and Merge Experiment

Experiment 22 explored plan division, improvement, and merging the workflow. First, a plan was divided into fragments. Then, the fragments were improved, and finally, after multiple redundant plan-fragments were generated, the best were selected and merged into an improved plan.

In a studio, work is traditionally distributed between architects by dividing the project based on spatial division or building systems. This work division method inspired the divide-improve-merge workflow.

Aims

The experiment aimed to evaluate a divide-improve-merge workflow through a comparison to an improvement task.

Figure 5.85: Improved Micro-task example screen (Experiment 22)

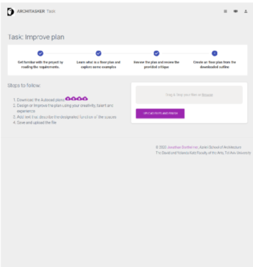


Figure 5.86: Improve plan micro-task steps (Experiment 22)

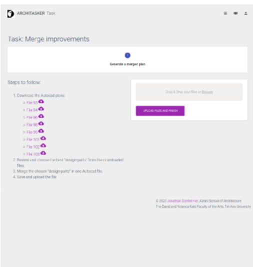


Figure 5.87: Merge micro-task steps (Experiment 22)

Method

Two new micro-tasks were developed — namely, 'Improve plan' and 'Merge fragments.' The 'Improve plan' micro-task consisted of the following four screens: Brief screen, Example screen (Figure 5.85), Base-artifact screen, and Task-steps screen (Figure 5.86). The task required the participants to download the AutoCAD file, improve and fix it, and then upload the new file to the software. The merge fragments task consisted of one task-steps screen (Figure 5.87). The task steps were as follows: downloading multiple AutoCAD files, reviewing them, choosing the best plan-fragments, merging the fragments into one plan, and uploading the new file to the software.

The experiment was conducted on April 30, 2019, with eight students. The participants were divided into two groups. The first group (six participants) was the test group. The second group (two participants) was the control group.

Each participant in the test group was assigned a plan improvement task with one plan-fragment. The task was to divide Artifact D43 into fragments, since this artifact, based on artifact D19, was made out of three separate structures. Each of the separate structures became one plan-fragment. The fragment AutoCAD file included the three structures and a polygon that outlined the fragment to be improved.

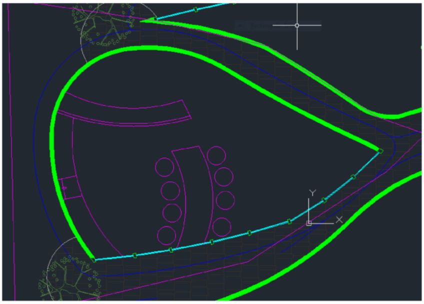
For each fragment, a functional program was provided: Fragment A (D47) was the dining area, Fragment B (D48) was the kitchen, and Fragment C (D49) was the toilets and storage. After the test group finished the 'Improve Plan' task, they were assigned the 'Merge Fragments' task.

In this task, the students were asked to select one from each fragment-set and to create a newly merged artifact.

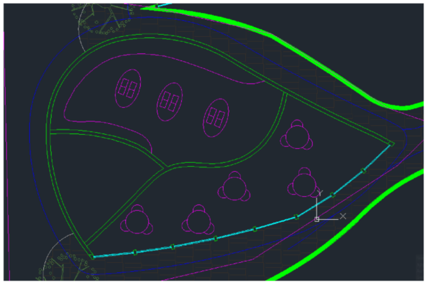
The control group was assigned an improvement task based on Artifact D43. They had to design the interior spaces using the same program. Upon completion of the tasks, the artifacts were analyzed, evaluated, and compared with a particular focus on the design quality.

Generated Designs and Data

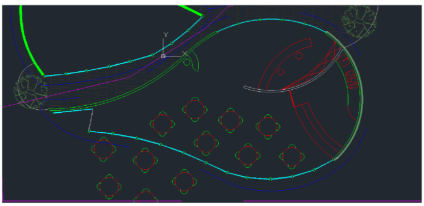
We expected the test group to generate six artifacts; however, only five plan-fragments were created (D50 - D55). One participant failed to provide a design due to time limitations. Therefore, for fragment C of the plan, there was only one artifact (D55).



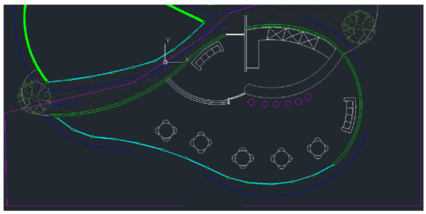
(a) Artifact D50 - Fragment A



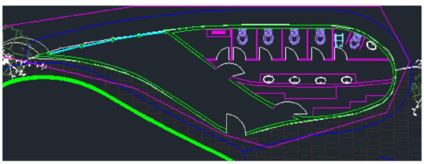
(b) Artifact D54 - Fragment A



(c) Artifact D51 - Fragment B



(d) Artifact D52 - Fragment B



(e) Artifact D55 - Fragment C

Figure 5.88: Fragments (Experiment 22)

The participants used these five artifacts and merged them, creating thus new artifacts. All fragments were used to generate the merged artifacts. However, the artifact-set D52-D50-D55 was perceived by the participants as the most favorable since most participants (N=4) chose these specific fragments.

The control group generated two artifacts: D56 and D63. Expert architects evaluated the merged artifacts and the control group artifacts. The relative rating is shown in Table 5.31.

Table 5.31: Merged artifacts and control group artifacts (Experiment 22)

Artifact	Set	Expert Evaluation
D57	D51-D54-D55	1
D58	D52-D54-D55	3
D59	D52-D50-D55	2
D60	D52-D50-D55	1
D61	D52-D50-D55	3
D62	D52-D50-D55	3
D56	Control	4
D63	Control	2

Analysis of Design and Data

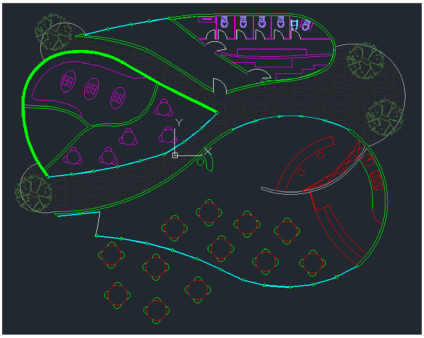
The results of our analysis of the designs are summarized below.

Based on fragment D47 (Fragment A):

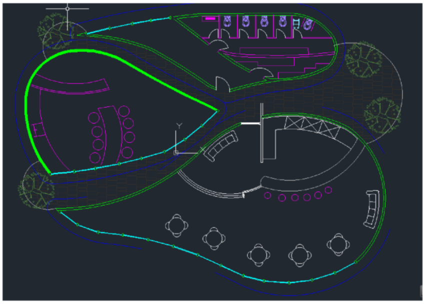
- D51: Added a kitchen, dining area, outdoor dining, and store.
- D52: Added an open kitchen with bar seating and dining area. It added a new entrance.

Based on fragment D48 (Fragment B):

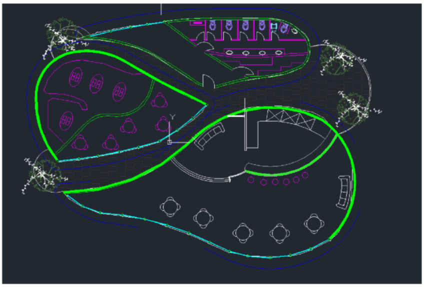
- D50: Added a kitchen with a counter and a bar table for eight people.
- D54: Divided the structure into three rooms (a kitchen, dining area, and an empty room) without connections.



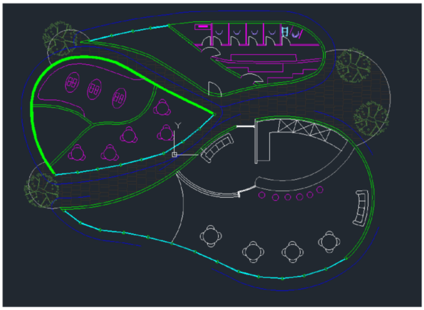
(a) Artifact D57 - made out of D51, D54 and D55



(b) Artifact D58 - made out of D52, D54 and D55

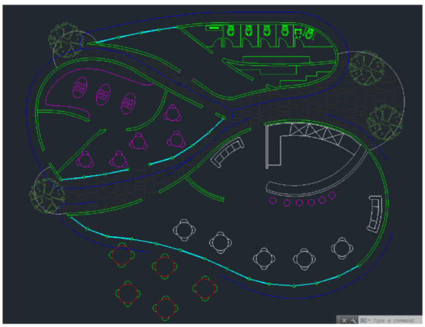


(c) Artifact D59 - made out of D52, D50 and D55



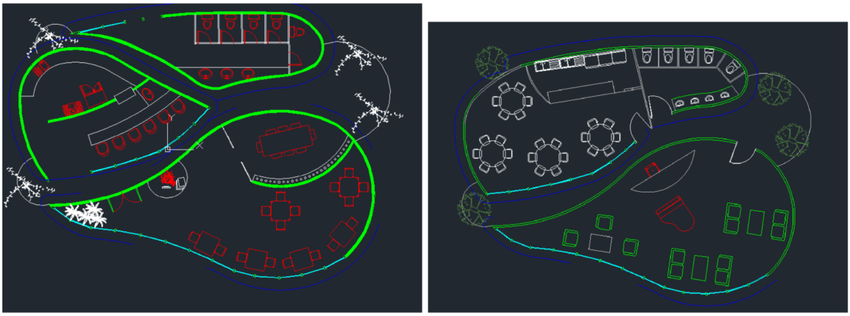
(d) Artifact D60 - made out of D52, D50 and D55

(e) Artifact D61 - made out of D52, D50 and D55



(f) Artifact D62 - made out of D52, D50 and D55

Figure 5.89: Merged artifacts (Experiment 22)



(a) Artifact D56 **(b)** Artifact D63

Figure 5.90: Control group artifacts (Experiment 22)

Based on fragment D49 (Fragment C):

- D55 had a washing room and storage room.

The above fragments were merged into 6 new artifacts:

- Artifact 57 was based on fragments 51, 54, and 55.
- Artifact 58 was based on fragments 52, 54, and 55.
- Artifacts 59, 60, 61 and 62 were based on fragments 52, 50, and 55.

The control group produced the following two artifacts:

- D56: Suggested a dining hall in the bottom part. The middle part had a bar and a kitchen. The top part had the toilets.
- D63: Merged the top and middle parts. The merged structure contained a kitchen, toilets, and a dining area. The bottom part included a sitting area.

Since there was only one option for fragment C and two options for fragments A and B, four artifact-set combinations were possible. The participants generated three set-combinations; however, one artifact-set combination was selected four times, highlighting the better fragments. Of note, the merged artifacts made from the same fragments were not identical and were evaluated separately.

On the other hand, the evaluation of the control group suggested that these students demonstrated a better design performance. Artifact D56 had the highest evaluation, which may be due to the fact that the control group had an overview of the entire structure and could better solve essential issues like circulation, entrances, and space function. For instance, Artifact D63 merged two structure parts, while D56 changed the locations of the doors, and both aforementioned artifacts targeted essential design issues.

While the experiment succeeded in designing and merging artifact fragments back to a merged artifact, the design quality identified in this process concerning the total design workflow was lower. This suggests that a divide-improve-merge workflow could be more useful at more advanced stages of the design in order to divide the work between crowd-workers and is less appropriate for the preliminary design stages.

Conclusions

The following conclusions were made:

1. The divide-improve-merge workflow succeeded in improving and generating merged artifacts.
2. The divide-improve-merge workflow may be used during advanced design and detailing stages.

5.4.7 Review and Artifact Improvement Experiment

Experiment 24 evaluated a new artifact improvement workflow that integrated two ideas from the previous review-improve experiments by requesting the participants to improve the whole plan and then merge the best improvements.

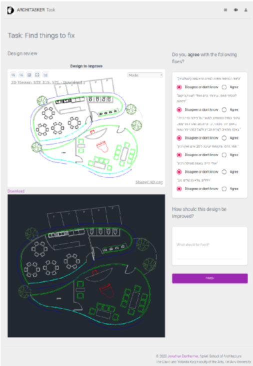


Figure 5.91: Find Things to Fix Screen (Experiment 24)

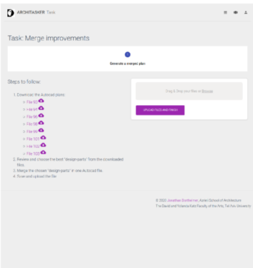


Figure 5.92: Merge Improvements Steps Screen (Experiment 24)

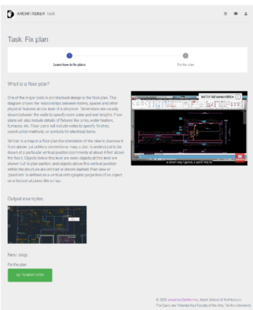


Figure 5.93: Fix Plan - Example Screen (Experiment 24)

Aims

The experiment aimed to evaluate a three-stage improvement workflow that included a review, artifact-improvement, and artifact-merge tasks.

Method

For this experiment, three new tasks were created. The first task, named 'Find Things to Fix' was similar to the review task, with only minor changes. The task contained only one question: 'How should this design be improved?' (see Figure5.91). In addition, an array of previously generated issues was displayed, allowing the participants to mark if the issues were resolved. If the participant marked the issue as unresolved, it was added to the participant's review. The requirement fulfillment form was omitted.

The second task was the 'Improve Artifact' task, which was made up of two screens. The first screen showed an example output (see Figure 5.93). In this task, the participants had to mark their improvements using a 'revision cloud,' an architectural graphic symbol to mark changes in plans. The second screen displayed the review issues and the task steps (see Figure 5.94). To conclude the task, an AutoCAD file had to be uploaded.

The third task, named 'Merge Improvement', consisted of a task-step screen (see Figure 5.92). The screen allowed the participants to download all previously improved AutoCAD plans, instructions on merging the 'best' improvements, and, finally,

The experiment was conducted on May 14, 2019, with six students. In the first task, which was based on Artifact D70, the participants generated review data that included design issues. The review task displayed previously generated review items. The participants had to mark whether or not the issue was resolved. If the issue was marked as not resolved, the issue was added as another review item.

The second task required the participants to download artifact D70 and improve it using the provided reviews. The improvements should be marked using a 'Revision cloud' marker, which is an accepted convention in architecture.

In the third task, the participants downloaded all the artifacts generated in the second task and, using the Revision Cloud marker, identified the plan's improvements. They had to choose the best artifact and merge the best improvements into it.

Finally, the artifacts and review data were evaluated and validated by expert architects.

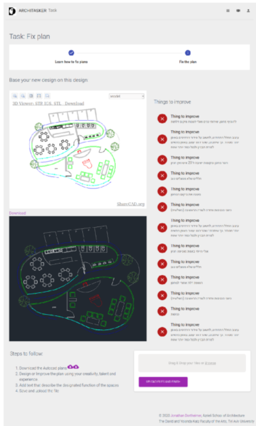


Figure 5.94: Fix Plan - Steps Screen (Experiment 24)

Generated Designs and Data

In the first task, a total of 21 review items were generated. The review item included previous reviews that were marked as unresolved.

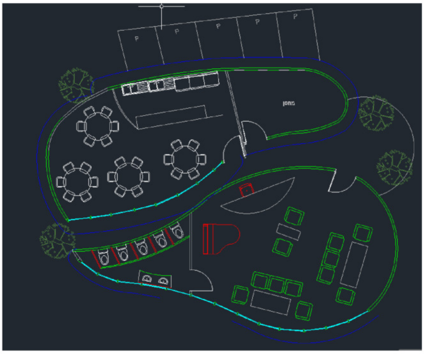
In the second task, the participants generated six new artifacts with revision cloud markers to indicate the improvements. The artifacts are presented in Table 5.32 and figure 5.95.

In the third task, each participant selected the artifact s/he thought was the best. Artifact D72 was selected three times, artifact D77 was selected twice, and artifact D74 was selected once. Then the participants merged improvements from other artifacts into their selected artifact. Six new artifacts were generated.

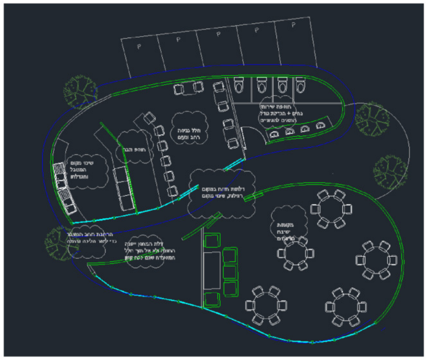
Analysis of Design and Data

The reviews were analyzed and categorized by topics. From the analysis, the following 15 topics were identified:

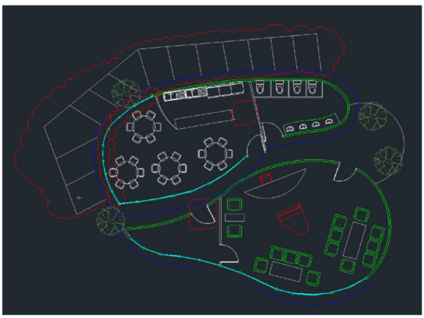
- R1 - Organizing sitting furniture - 7 reviews
- R2 - Entrances locations - 6 reviews
- R3 - Return to previous form - 3 reviews
- R4 - Car parking - 2 reviews



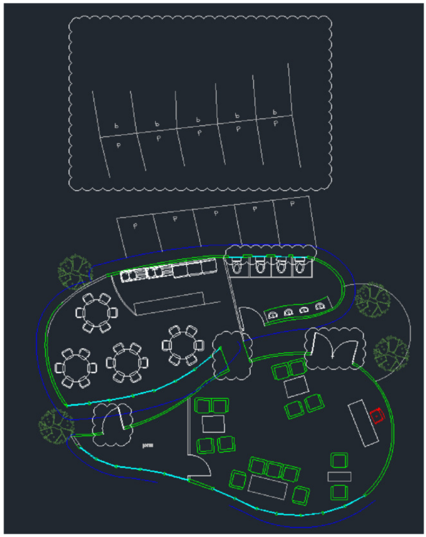
(a) Artifact D71



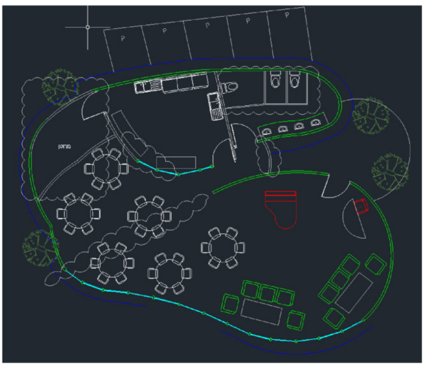
(b) Artifact D72



(c) Artifact D73



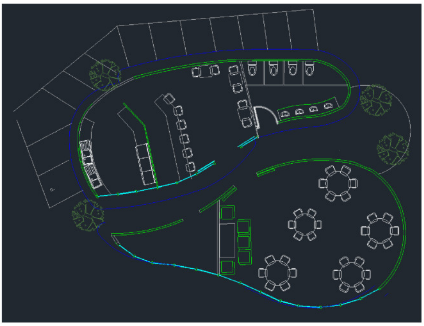
(d) Artifact D74



(e) Artifact D76

(f) Artifact D77

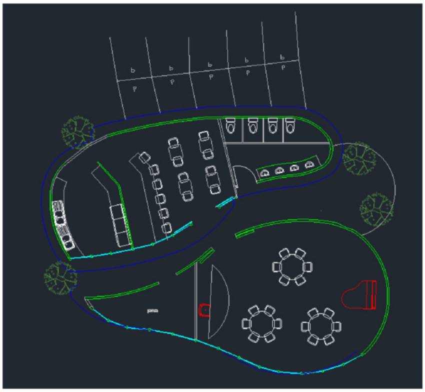
Figure 5.85: Layered artifacts (Experiment 24)



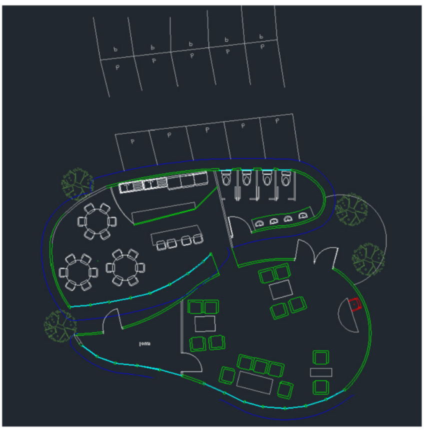
(a) Artifact D78

(b) Artifact D79

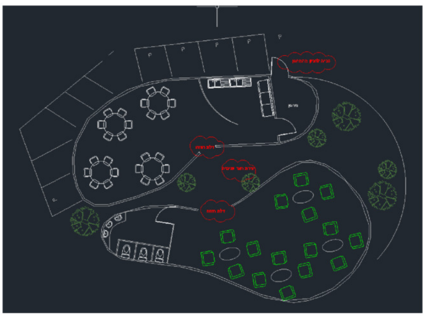
(c) Artifact D80



(d) Artifact D81



(e) Artifact D82



(f) Artifact D83

Figure 5.96: Merged artifacts (Experiment 24)

- R5 - Spaces are not used well - 2 reviews
- R6 - Kitchen organization - 1 review
- R7 - Optimizing courtyard - 1 review
- R8 - Storage needs to be near the kitchen - 1 review
- R9 - Sitting area needs to be near to the dining area - 1 review
- R10 - Storage does not require curtain walls - 1 review
- R11 - WC required windows - 1 review
- R12 - Replace WC with storage - 1 review
- R13 - Add accessible WC - 1 review
- R14 - Kitchen too small - 1 review
- R15 - Add a Bar - 1 review

The issues that emerged in the review data indicated some design problems that were also identified by the experts. For example, the most problematic issue was the disconnection of functions since the building was divided into separated structures. Another problematic issue was circulation, as the entrance of the building needed to be outbound facing but also connect the structures.

However, some reviews were conflicting. For example, changing the building from two structures to a three structure building would have resulted in more circulation and functionality issues. Since this solution was not practical, all participants ignored it. The analysis showed that providing past reviews generated more reviews that were less relevant to the current design. Design issues in the new artifacts were already mentioned in the new reviews.

The analysis of the generated artifacts showed which review topics were resolved (see Table 5.32. Since some artifacts were improved, not all review issues could be solved. For example, R8 and R12 were conflicting since both suggested different ideas regarding the storage location.

In the second task, the participants had to select the artifact they liked the most. As mentioned above, Artifact D72 was selected three times, Artifact D77 was selected twice, and Artifact D74 was selected once. While Artifact

Table 5.32: Artifacts and review resolve rate

Artifact	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	Resolve rate
D71	Yes	No	No	Yes	No	No	No	Yes	No	Yes	Yes	Yes	No	No	No	40%
D72	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	73%
D73	No	Yes	No	Yes	No	No	No	No	No	No	No	No	No	No	No	13%
D74	Yes	Yes	No	Yes	No	No	No	No	Yes	No	Yes	No	No	No	No	33%
D76	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	No	73%
D77	Yes	Yes	No	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	No	66%

D74 had a low resolve rate, it was selected exclusively by its author, highlighting that, in the absence of filtering in a selection task, there is a risk of bias and waste of resources. On the other hand, Artifacts D77 and D72 were evaluated by an expert architect as better design solutions. The artifacts selection outcome is presented in Figure 5.97.

The results showed that the artifacts' selection was successful in choosing the artifacts that solved most of the issues since the participants deeply involved in the task were also selecting the artifacts and had prior knowledge to identify the best solutions.

In the third task, the participants reviewed all generated artifacts. They were asked to pick the best design solutions from the artifacts and to copy them into the artifact of their choice. Six artifacts were generated and reanalyzed (see Table 5.33). Since the new artifacts were based on more successful artifacts, and due to the removal of the artifacts with low resolve rates, the average issue resolve rate increased from 50% in the second task to about 68.89% in the third task.

Table 5.33: Merged artifacts and review resolve rate (Experiment 24)

Artifact	Based on	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	Resolve rate
D78	D72	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	73%
D79	D72	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	73%
D80	D77	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	80%
D81	D72	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	73%
D82	D74	No	Yes	No	Yes	No	No	No	No	Yes	No	Yes	No	No	No	Yes	33%
D83	D77	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	80%

Furthermore, Artifacts D78, D79, and D81 did not improve the resolution rate, while they did had design improvements. Artifact D82 was based on artifact D74 and was further developed by the same designer. However, it did not show any improvement concerning resolving the reviews.

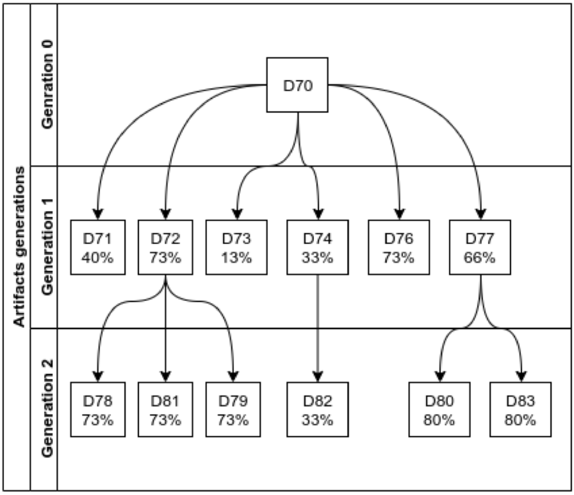


Figure 5.97: Artifact selection outcome (Experiment 24)

Although most artifacts fixed the review issues, the expert opinion suggested that the artifacts were not sufficiently high quality. The fixes were evaluated by the experts as making only small improvements instead of viewing the comprehensive design. Admittedly, the resulted artifact design was still problematic and had some significant circulation and unfixed programmatic issues. The challenge in fixing these issues may be related to the challenging use of AutoCAD with non-orthogonal forms, which requires a high degree of expertise. This issue will be addressed in the next experiment.

Conclusions

The following conclusions were made:

1. The improvement task was successful in fixing the provided reviews.
2. Most participants selected the artifacts that resolved most review issues. Therefore, the selection mechanism can effectively identify good designs.

3. The unmanaged selection process resulted in the selection of better artifacts and some sporadic selection of less successful artifacts. Therefore, a threshold should be introduced that would help to remove the selection of sporadic artifacts.
4. Not all review issues could be resolved, as some issues were conflicting. Therefore, we suggest adding another stage to organize and aggregate the reviews.

5.4.8 Improvement Through Sketching

Experiment 25 explored the use of sketching as a technique for the artifact improvement task.

Aims

The experiment aimed to evaluate sketching, which is widely used in architecture, as a problem-solving method [107]. We expected to observe significant improvements.

[107]: Goldschmidt (1992), "Serial sketching: visual problem solving in designing"

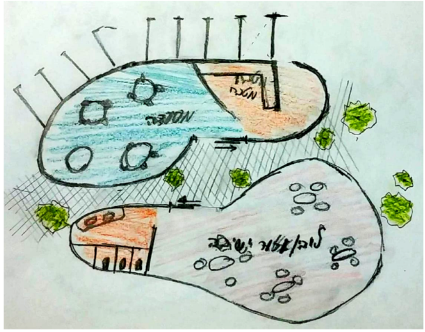
Method

The experiment was conducted on May 21, 2019, with eight students. Each participant was assigned a sketch generation task that consisted of a 'Brief screen,' 'Example screen,' 'Base-artifact screen,' and a 'Task-steps screen.' The participants were required to create a programmatic sketch and solve the design issues that emerged in previous experiments. The participants were provided with sketching paper and six printed plan-artifacts (Artifacts D78 -D83).

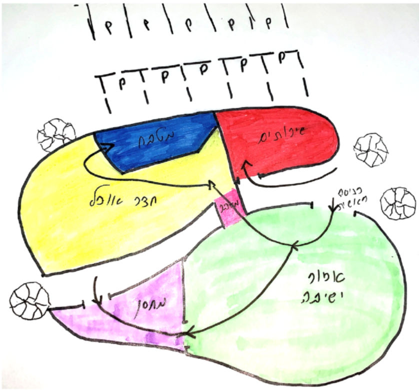
Generated Designs and Data

The task generated the following eight schematic sketches (see Figures **??** and **??**):

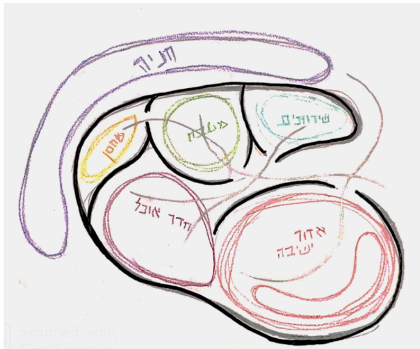
(a) Artifact D84



(b) Artifact D85



(c) Artifact D87



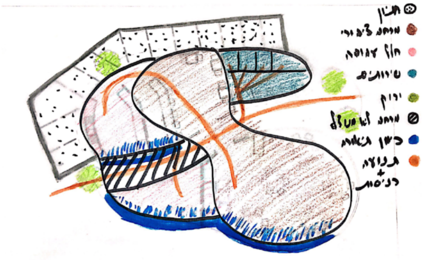
(d) Artifact D93

Figure 5.98: Artifacts (Experiment 25)

- D97 based on D78
- D94 and D96 based on D79
- D84 based on D81
- D93 based on D80
- D85 and D95 based on D83
- D87 based on D82

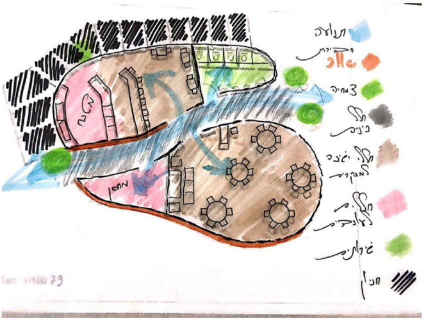
Analysis of Design and Data

We examined the new artifacts. Six artifacts (D84, D87, D85, D95 D96, and D97) were colored plans without changes or improvements. These schemas copied the

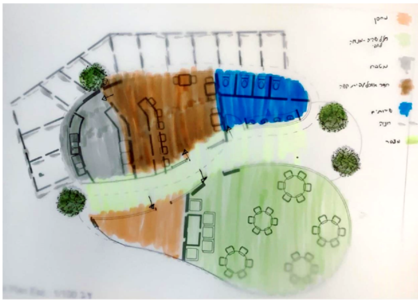


(a) Artifact D94

(b) Artifact D95



(c) Artifact D96



(d) Artifact D97

Figure 5.99: Artifacts (Experiment 25)

exact situation from the base-artifact without presenting new solutions.

However, two artifacts introduced the following improvements to the design:

- D93 was an abstract schema displaying a flow between the spaces and functions. The schema offered a better connection between the different spaces and functions.
- D94 provided a spatial solution where the two structures were connected. The schema changed the entrance location by moving it to the middle of the structure, which made sense. The new schema continued the curvy lines of the design.

Based on the results, we concluded that the task requirements were not sufficiently clear, and most participants focused more on the graphic task than on problem-solving.

However, the task produced two new artifacts that resolved the previously unresolved issues. This was a partial success since we considered these artifacts as better than the solutions obtained in Experiments 23 and 24 that used AutoCAD.

Conclusions

The following conclusions were made:

- Sketching is more effective than using CAD for conceptual design problems.
- Further research is needed to optimize the task instructions.

5.4.9 Plan to Concept Sketch

Experiment 26 examined the process of transforming a 2D plan into a 3D sketch. This process is essential, as the building plans were significantly changed and had notable changes that affected the structure's form.

Aims

The aims of the experiment were to 1) evaluate a task of developing a 3D conceptual sketch artifact from a 2D plan artifact; and 2) find out the preferable architectural medium for this stage by comparing 3D CAD modeling and conceptual sketching.

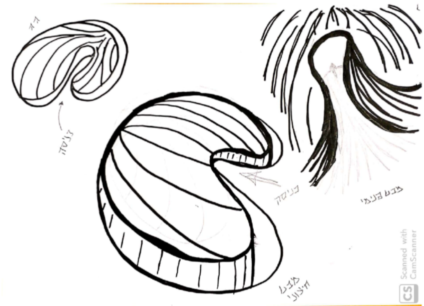
Method

The experiment was performed on May 21, 2019, with eight students. The participants were asked to generate a 3D sketch based on the previously selected Artifact D93. We allowed the students to choose sketching or

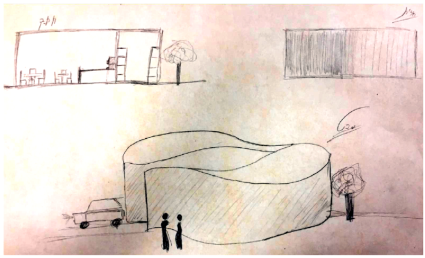
CAD modeling to create the new artifact. Finally, the new artifacts were collected and analyzed.

Generated Designs and Data

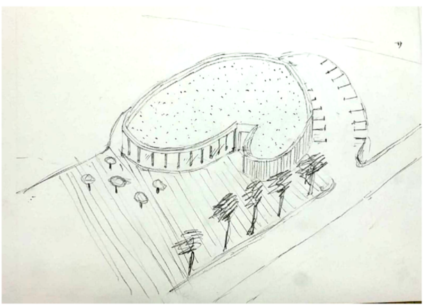
The participants generated eight new artifacts (Figure 5.100 and 5.101 and Table 5.34.



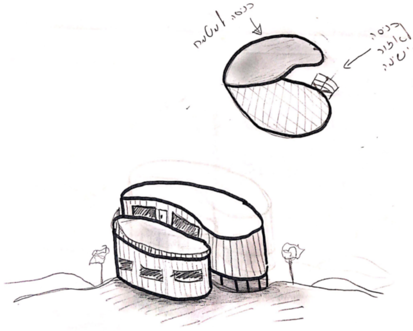
(a) Artifact D98



(b) Artifact D99



(c) Artifact D100

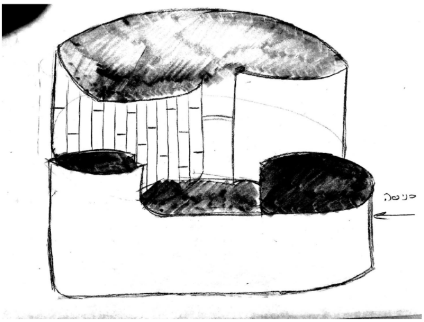


(d) Artifact D101

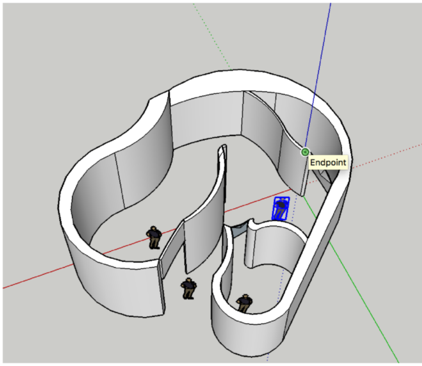
Figure 5.100: Artifacts (Experiment 26)

Analysis of Design and Data

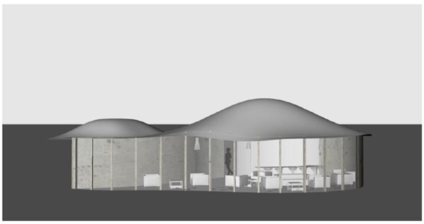
Eight artifacts were produced, of which six artifacts were sketches, and two were CAD 3D models. The results of expert evaluation of the sketches were as follows:



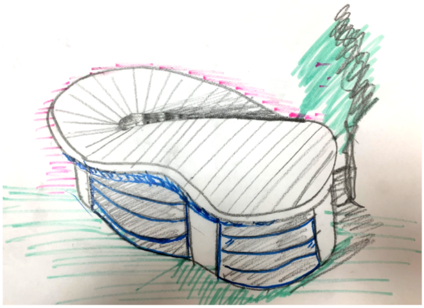
(a) Artifact D102



(b) Artifact D103



(c) Artifact D104



(d) Artifact D105

Figure 5.101: Artifacts (Experiment 26)

- D98 was a sketch of an ameba form structure with a sloped roof and closed with glass facades.
- D99 was a sketch of an extruded form covered with a wave roof. However, the elevations and sections were rectangular.
- D100 was a sketch of an extruded curved structure covered with a flat roof. The structure also included a terrace, and a colonnade supported its roof.
- D101 was a sketch of a structure made of two ameba-formed structure wings of different heights. A flat roof covered both, and the structure had rectangular windows. One of the wings was positioned on the ground, while the other was supported by small pillars.

Artifact	Experience Years	Method	Expert Evaluation
D98	3	Sketch	2.5
D99	3	Sketch	2.2
D100	2	Sketch	2.2
D101	3	Sketch	1.4
D102	3	Sketch	1.1
D103	2	CAD	3.9
D104	4	CAD	4.2
D105	2	Sketch	2.5

Table 5.34: Artifacts, participant's experience, and expert evaluation (Experiment 26)

- D102 was a sketch of a two-wing structure, similarly to D101.
- D105 was a sketch of a curved structure with a sloped roof and curtain walls.

The results of expert evaluation of the CAD artifacts were as follows:

- D103 was a CAD model that extruded the base artifact walls and showed the interior details of the design.
- D104 was a CAD model of a structure with glazed curtain walls. It was covered by a roof raised in the form centers.

Overall, the CAD artifacts were evaluated more positively than the sketches. The experts commented that the higher evaluation was a result of clarity of the architectural representation of the CAD artifacts.

The task was successful in transforming the plan artifacts into 3D representations. The results highlight the advantages of 3D modeling using CAD software for these phases.

Conclusions

The following conclusions were made:

1. The task successfully generated 3D architectural representations.
2. At this stage, CAD models are a better choice for the design process due to their higher clarity and ability to provide richer detail.

Discussion 6

The discussion chapter starts with reiterating the main research question by discussing the architectural crowdsourcing model's structure and operation. We introduce our model on three levels and present the DSR block concept as a generic micro-competition component for design crowdsourcing (see Section 6.1).

Next, we discuss the secondary research questions dealing with the structure and GUI of the design, selection and review micro-tasks (see Sections 6.1.4, 6.1.5, and 6.1.6). Then, we present the results with regard to the use of different design tools in the crowdsourcing process (see Section 6.2).

Upon the discussion of the crowdsourcing model, we reflect on the literature and discuss theoretical implications. First, we analyze our findings in the light of design methods research (Section 6.3). We also discuss several important quality considerations (Section 6.4). Upon a review of the implications of the process as a participatory design process (Section 6.5), we formulate several recommendations for architecture competitions and practice (see Section 6.6).

Finally, we propose a new notion of open-source architecture (Section 6.7) and we acknowledge the limitations of the present study and outline future research directions (Section 6.8).

6.1 Crowdsourcing Model for Architectural Design

Regarding the main research questions addressed in this study (*"What kind of crowdsourcing workflow and micro-tasks are preferable in architectural design to solve the design requirements, provide higher design quality, and is easier to use according to the participants' and expert architects' opinions?"*), the results of the first set of experiments in

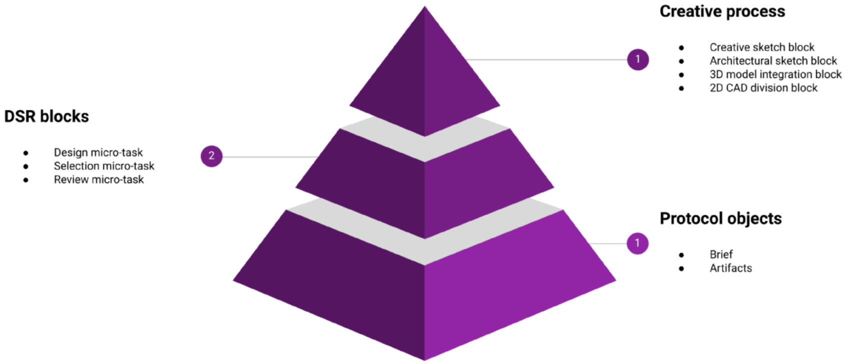


Figure 6.1: The levels of the Crowdsourcing Process model for Architectural design. Protocol objects, DSR block and high-level design process

workshop 1 (Projects 1, 2, and 3) provided numerous insights about the design process, user interface, and the application of various design tools. We formulated an improved workflow, which was consequently applied as a new software. In the second set of experiments (Project 4), we examined the application of the improved design process and investigated the issues related to the development of the artifacts.

To this end, we first introduce a new architectural design crowdsourcing model described in three levels. First, the Protocol level includes objects used to communicate between the higher-level components. The DSR block level is the second level and is a generic micro-competition component. The design process is the third level and is made out of different and multiple DSR blocks.

6.1.1 Protocol Objects

The information transferred between micro-tasks is defined by the following two types of objects: (1) the brief object and (2) the artifact objects.

The *brief object* is a data structure loosely based on RIBA's plan of work [63]. Since an architectural project brief

[63]: RIBA (2013), *RIBA Plan of Work 2013*

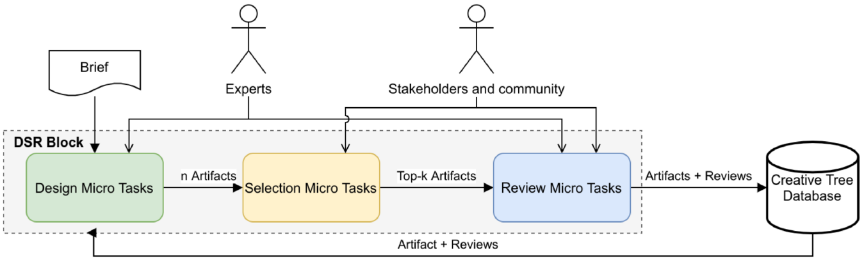


Figure 6.2: Design-Select-Review (DSR) block process, with participant types.

might span hundreds of pages, which makes it impractical to use it for micro-tasks that are only several hours long, the brief object is a summary of a standard brief and includes the project name, desired outcomes, client profile, project objectives, project stakeholders, and links to various resources such as Wikipedia pages, Google Earth, or photos of the area and precedents. The brief also includes essential site information, such as a site CAD model, which includes the plot and surroundings to provide a better understanding of the existing spatial settings.

Artifact objects, which hold a reference to traditional architectural artifacts by having graphic or CAD files, contain textual descriptions and review information to facilitate the improvement process. The artifact objects, which are generated in design micro-tasks, later become inputs for the next micro-tasks that build upon them. Our model includes the following artifacts: concept sketches, architectural sketches, 2D CAD drawings, and 3D CAD models.

6.1.2 DSR Block

We introduce the term and concept of DSR blocks as building blocks of design-crowdsourcing workflows. A DSR block consists of the following three sets of tasks: (1) Design tasks; (2) Selection tasks; and (3) Review tasks (see Figure 6.2). The first letters in the names of these three types of tasks make up the name of a block—i.e., DSR. According to the type of artifact produced in the tasks, the following four kinds of blocks were developed: (1) 3D conceptual sketch; (2) 2D architectural sketch; (3) 3D digital model; and (4) 2D digital drawing.

6.1.3 Design Process

As discussed in Chapter 4, we hypothesized that the design process was iterative and based on design, selection, and review tasks. The improved process presented in the present dissertation reinstates the iterative concept but is more complex and detailed.

In Experiment 1, the use of manual sketches proved to be an effective means of displaying architectural ideas within a short time.

We experimented with the transition between sketching and CAD and between 2D drawings and 3D models in further experiments. Switching between the different architectural mediums rendered the process more complex. In order to obtain a structured outline of this process, we defined Design-Select-Review blocks on the lower level of the design process.

As mentioned previously, the design process consists of several DSR blocks (see Figure 6.3). First, a conceptual sketch block generates multiple architectural ideas as a three-dimensional perspective (Experiments 1, 8, and 16). The subsequent block develops these conceptual sketches to two-dimensional architectural sketches: plans, elevations, and sections (see Experiments 10 and 25).

The next block integrates the conceptual sketch with the new architectural sketches (plans, elevations, and sections) into a digital three-dimensional accurate model (see Experiments 4, 11, and 19). Since the sketches are produced by different participants and are often inconsistent, they do not fully match each other. As the discrepancies among different sketches are not yet resolved, designers must make decisions and resolve them. This results in the establishment of an accurate and holistic CAD model that may serve as an outcome for the crowdsourcing process.

In this stage, it is possible to stop the design process since information-rich 3D CAD models are generated and selected. The design process stops by reaching the maximum defined number of iterations or upon a project manager's decision. The decision is based on the manager's judgment that takes into account the generated artifacts design fitness, quality, budget, and time constraints.

However, in the early stages, the three-dimensional model is still preliminary, so there is a need to improve the design through the following DSR blocks.

In Experiments 7 and 13, we hypothesized that it would be possible to facilitate a design improvement process by editing and improving the SketchUp models. However, contrary to the expectation, the improved models did not advance the design and, for the most part, deteriorated the design. Nevertheless, in Experiment 14 that explored the design's improvement through plans, we found that the change in the medium allowed for providing more details with a two-dimensional drawing.

Therefore, the next block transformed the model into architectural CAD drawings—including plans, elevations, and sections. As shown by the results of Experiments 20 and 21, making "section-planes" in this three-dimensional model was too complicated for the participants without prior knowledge of operating SketchUp and AutoCAD. However, the results of Experiments 21, 22, and 23 demonstrated that the same participants could develop the design after plans were created from the three-dimensional model.

Upon the production of plans, sections, and elevations, a significant challenge emerged to improve them to facilitate an iterative design process. In Chapter 5, we reported the results of several experiments performed to improve the design. For instance, Experiment 22 examined the improvement of CAD plans by fragmenting them. Different participants improved the fragments, which was followed

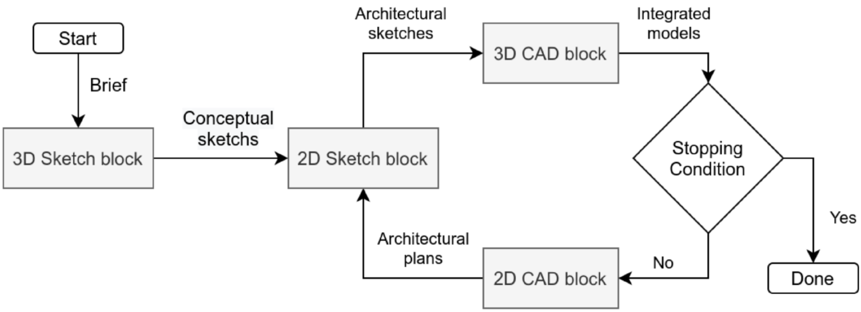


Figure 6.3: Architectural DSR Block Crowdsourcing Process

by a consolidation of the produced fragments by selecting the most successful parts and creating merged plans. In this experiment, the control group that improved the whole plan was more successful in terms of providing higher evaluated designs.

Furthermore, Experiments 23 and 24 examined the improvement plans based on the review data provided in the review micro-task, cherry-picking the improved plans, and merging the improvements into new plans. The results of these two experiments showed that, although the participants focused on improving issues from the reviews, the designs still had significant unresolved issues like the disorganization of the functions, problematic entrance to the building, and more. As in Experiment 22, the participants in Experiments 23 and 24 selected the best parts from the plans and generated new merged plans. However, a necessary holistic approach to solving design problems was not observed yet.

However, Experiment 25 was successful in improving the designs using sketching instead of CAD. The transition from the rigid and precise medium of the AutoCAD software to the freedom of the pencil showed that using sketches to solve fundamental design problems in the preliminary stages of the design process was more successful than CAD. Therefore, transforming CAD back to architectural sketches closed the loop in the design process.

Overall, the experimental results revealed that the proposed design process is circular. In the following steps, the architectural sketches were merged again into a three-dimensional digital model, and so on. These results support the presented model that provides a conclusive answer to our main research question [Main-RQ].

6.1.4 Design Micro-Task

The first secondary research question [RQ-1] asked, *"Which type of micro-task yields artifacts that are evaluated higher by experts?"*. Based on the experience gained during the experiments, we propose a design task type as implemented in recent experiments. Specifically, the developed design task consists of 3 to 4 screens detailed below.

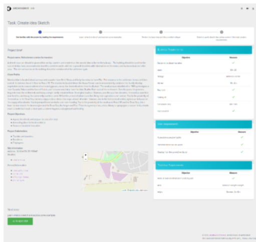


Figure 6.4: Design Micro-Task Brief Screen

The first screen presents the project brief (see Figure 6.4). Typically, an architectural project brief is a multi-page document that includes detailed requirements and many technicalities. However, for a micro-task with stringent time limitations, such a document is not suitable. Therefore, the crowdsourcing brief used in our experiments included only essential information, such as design goals, client information, stakeholder analysis, climate, geography, history, as well as the most critical requirements for business, users and technical aspects. Besides, the brief must also include a map of the site and a 3D digital model with a marking of the construction site to be used as a foundation for the artifacts.

Figure 6.5: Design Micro-Task Example Screen

The second screen presents an example of the desired outcome to clarify the expectations of the designer (see Figure 6.5). The screen contains a verbal description explaining the importance of the type of artifact that needs to be generated, along with a video demonstrating how the task should be performed. Finally, several images with examples of deliverables are provided. The example screen is necessary to coordinate expectations with the designers. On the one hand, the examples express an expectation of the required architectural quality. On the other hand, the examples establish a work frame so that

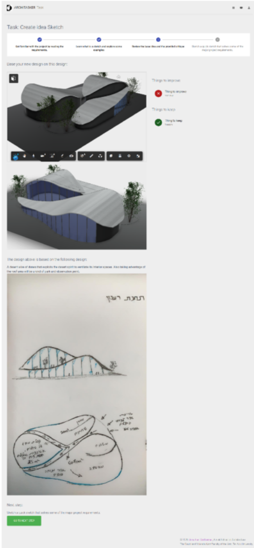


Figure 6.6: Design Micro-Task Artifact Screen

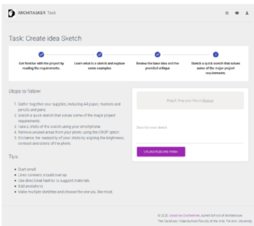


Figure 6.7: Steps Screen

designers do not invest too much time in details that are not required.

The third screen exhibits the artifact needed to be developed on the current task (see Figure 6.6). This screen does not appear in conceptual sketch tasks since artifacts have not been created yet. The screen includes images and dedicated viewers to navigate 3D and 2D CAD files. Next to the artifact, there is a list of previously created review items and quantitative requirements scores. This information helps designers to quickly understand the gaps and shortcomings of the current artifact. As shown by the results from Experiments 7, 23, and 24, most designers addressed the reviews and improved the artifacts according to both review items and quantitative requirements scores. The goal of providing artifacts from the previous tasks is to provide a broader context of the design process for the designers.

The fourth and last screen is the task execution screen that contains a list of simple steps towards the task performance (see Figure 6.7). The results of Experiment 8 highlighted the importance of simple steps. The design stages of the various artifact design tasks are quite similar and, in the end, the designers send the files of the new artifacts to the software. Furthermore, the results of Experiment 1 emphasized the importance of clear photographing instructions, so that manual sketches are presented well. In addition, as demonstrated in Experiment 8, it is also essential to limit the number of sketches that a designer can produce - the more sketches - the lower their quality.

Finally, as suggested by the results of Experiment 20, design tasks should be simple and include a limited number of steps. According to our observations, complex requirements led to the participants' failure to produce artifacts. Another essential aspect is using cloud file storage services, such as the Trimble Connect[1] or Autodesk A360[2], for displaying 3D models in software; however, using these cloud services is complicated, confusing, and requires very explicit instructions. Based on the results of Experiments 11 and 19, we would recommend integrating

1: An AEC collaboration platform by Trimble. See https://connect.trimble.com/

2: A design collaboration platform by AutoDesk. See https://a360.autodesk.com/

these services and software to facilitate the participants' performance.

6.1.5 Selection Micro-Task

The second secondary research question [RQ-2] was *"Which micro-task yields an artifact selection that is closer to expert evaluation?"*. While the design micro-tasks create various design alternatives proposing different approaches, a useful technique to choose the most suitable artifacts is required. Therefore, the selection task is a critical part of the design process and outlines the project's creative path.

In the first experiments, we explored selection tasks that included rating artifacts based on different criteria, similar to the techniques identified in the literature review [98]. In such rating tasks, the participants are presented with a single artifact and are asked to rate it using several criteria using a scale. For instance, in Experiment 2, the participants were asked to rate the presented artifacts according to their perceived quality, innovation, and functionality. Furthermore, in Experiment 5, the participants rated the artifacts according to the various metrics from the Safra Square competition protocol (see Pilot experiment). In addition, in Experiment 9, the participants rated the artifacts according to their perceived design quality, idea, compliance with requirements, as well as whether they would choose a given artifact to design down the road.

[98]: Wu et al. (2015), "An evaluation methodology for crowdsourced design"

The results showed that parameters such as functionality and stability did not correlate with the expert evaluation rating. On the other hand, the results revealed a correlation between expert evaluation and design quality. We conclude that our expert architects evaluated the artifacts' total quality based on the artifacts' aesthetic parameters. This correlation also indicates that the students and experts converged in their evaluation of the design quality of the artifacts.

In Experiment 11, we examined a selection method that was based on a comparison of two artifacts. According to the results, while the selection task failed to highlight the most suitable artifacts, we found that the participants liked the selection interface. To pursue this insight, in Experiment 15, we examined a sets-creation task based on a selection user interface. The results showed a clear choice of higher-quality artifacts, suggesting a breakthrough in the following three crucial aspects: (1) the selection of artifacts was more straightforward for the participants; (2) the selection produced more distinct results as compared to rating results; and (3) the task required less time. In Experiment 18, we validated the association between the artifact selection task results and expert evaluation.

In the results of the first selection experiments, we observed a bias in designers' performance on the selection tasks. Since this bias could have been caused by the inclusion of the designers' own artifacts in the task, in all subsequent rating task experiments, we prevented the designers from rating their own producted artifacts. However, selecting the most suitable artifact was more problematic since omitting one artifact from the list of possible artifacts could distort the results. Therefore, we conclude that, in further research, it is vital to find a method that would account for a likely author's bias. Our study found that reducing the weight of the value of an author's vote for their artifact helped reduce the bias effect and allowed the participants to select the most suitable artifacts more easily.

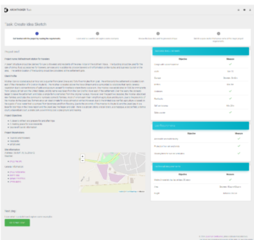


Figure 6.8: Select Micro-Task Brief Screen

The final developed selection task was based on two screens (see Figure 6.8 and 6.8). The first screen was the brief screen. The following screens exhibited all artifacts, and the participants were required to choose the artifact with the most promising design by clicking on it. The 3D artifacts were displayed using the Autodesk A360 viewer that allows for viewing the model from different directions, as well as to zoom in and out as needed.

Overall, the participants' selection results can be presented as a histogram of artifact choices. The unselected artifacts are typically eliminated from the process. Furthermore, based on the results of Experiment 5, we concluded that some of the less frequently selected artifacts should be removed so that at least two artifacts remain. This will allow for the development of competing idea branches [87] and choosing more suitable designs using open-source development methods [2].

[87]: Sun et al. (2015), "Collaborative sketching in crowdsourcing design: a new method for idea generation"

[2]: Carpo (2011), *The Alphabet and the Algorithm*

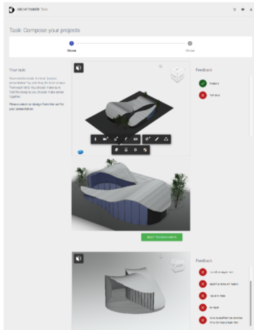


Figure 6.9: Select Micro-Task Example Screen

In addition, the results of Experiments 10, 13, and 24 showed that whenever designers choose artifacts for further development as part of a design task, their choice is well aligned with the experts' opinion. Therefore, the designers' judgment is incorporated into the selection process, and they are granted creative freedom. On the other hand, due to the bias mentioned above, the designers' choice should not be the only criterion to rely on. Based on these considerations, we concluded that it is not effective to leave the selection of artifacts for further development exclusively to the designers, as it is inefficient to invest work into the artifacts that will definitely be removed afterward.

A collaborative design process based on a team or community should allow for diverse opinions, and this is one of the advantages of the proposed method. However, in the end, the design process needs to flow into a specific artifact. In the advanced iterations of the design process, artifacts are derived from the first generation. The design process can be viewed as a tree. The branches are narrowed until old branches are omitted, as also suggested by Sun et al. (2015). However, new artifacts are always generated as "leaves" on a branch of the design tree. Therefore, throughout the process, the designs are depleted of a specific design idea, as old idea branches become obsolete and new ideas emerge. There is a fine line between multiplicity and singularity of the design tree that affects the dynamics of the design process.

To summarize, the developed selection task provided the answer to the second research question [RQ-2] and helped to simply and effectively manage the design process in selecting the most successful design options.

6.1.6 Review Micro-Task

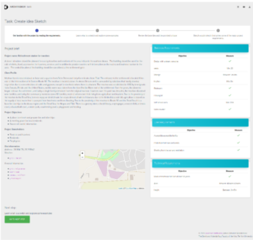


Figure 6.10: Review Micro-Task Brief Screen

Figure 6.11: Review Micro-Task Example Screen

A design process improves and develops conceptual ideas to established designs. Theoretically, any design should evolve and solve problems that arise from the initial design. The review task was designed to collect design issues from the participants. During the experiments, we examined different types of review tasks, such as free-text input and well-structured input to address the question [RQ-3] (*"Which type of micro-task yields design reviews that are beneficial to designers?"*),

The results of Experiment 9 showed that the participants found the review of too many artifacts to be exhausting. Therefore, we concluded that, in order to limit the number of reviewed artifacts, the review task should follow a selection task. Since providing reviews is a time-consuming task, it does not make sense to review artifacts removed in the selection task right after. Therefore, the review task follows the selection task.

Furthermore, the results of the experiments revealed that the higher the number of review questions, the shorter and less detailed the reviews. Therefore, we concluded that a review task could include only one question to be answered (e.g., "How would you improve this design?").

In Experiments 3 and 6, reviews previously given to the current artifact by the rest of the participants were presented on the review task screen. This was helpful by suggesting examples reviews and reducing the number of duplicate reviews. These results highlight the importance of providing the participants with relevant examples and providing an overview of the collected reviews.

The final developed review task was based on two screens (see Figure 6.10 and 6.11). The first screen was the brief screen. The following screen displayed an artifact with an input form that allows providing an answer to the presented question.

6.2 Design Crowdsourcing Tools

The second secondary research question [RQ-4] was *"Which design tools are suitable for the various stages in the crowdsourcing workflow?"* . The hypothesis was that a crowdsourcing process would be based on using the SketchUp software as a design tool. With the onset of experiments and aiming to deal with the inexperience of the participants with the software, we started with sketches. The results showed that, due to simplicity and the possibility to finish them within a short time, sketches were most useful for design micro-tasks. Another factor that contributed to the participants' creation of sketches was that the study participants had at least 1-year experience in architecture studies.

Moreover, working with SketchUp was difficult for the participants who had limited CAD experience (mainly with AutoCAD and Rhino). Despite the intuitiveness of the SketchUp software and the effort we put into training, the students experienced difficulties expressing their ideas (see Experiments 4, 7, and 12) and improving the artifacts (Experiment 13). There was mainly the difficulty in expressing complex geometries and curved lines. In contrast, freelance architects recruited from Upwork were able to produce high-quality and elaborate artifacts (see Experiment 19). This indicates that knowledge and experience are essential for the use of professional design tools.

On the other hand, the use of AutoCAD software was more successful. Similarly to working with sketches, we observed that the successful results had to do with the students' previous knowledge of and experience in using the software (see Experiment 14). Still, we encountered difficulties in using the software to improve plans in Experiments 20-24. Based on the results, we concluded that, although the participants had some experience with AutoCAD, they were not sufficiently experienced to perform complex design tasks such as revising and solving design problems.

However, an alternative possibility is that the participants with more experience with AutoCAD and SketchUp would have more successfully improved the artifacts using these tools. This possibility requires further investigation in future studies.

6.3 Design Process Theory

The concept of the DSR blocks is similar to the notions from previous research based on ideas of "solution space" search and exploration [5, 58], rather than fragmentation of the problem [86]. Our model is similar to the existing competition model [46], since both explore the solution space and help to select the best solutions.

[5]: Simon (1969), *The Sciences of the Artificial*
[58]: Corne et al. (1994), "Solving design problems by computational exploration"
[86]: Kittur et al. (2011), "CrowdForge Crowdsourcing Complex Work"
[46]: Angelico et al. (2012), "Crowdsourcing Architecture : a Disruptive Model in Architectural Practice"

However, the DSR blocks have some benefits over competition. First, instead of fully developed competition entries, a DSR block includes only micro-tasks with a limited scope, which allows more control over the design process, as intervention is possible between design iterations. The second advantage of the DSR blocks is that each design micro-task requires a small time investment and allows a small but fair reward to the designers. Third, since the DSR blocks are repeated multiple times and multiple designers develop the artifacts, the design product is an outcome of a collaborative design process, facilitating collective intelligence.

In what follows, we discuss the crowdsourcing model with regard to the design methods theories presented in Chapter 2. Our focus is on the following two aspects: functional and structural. While the functional aspect defines the components that are part of the design process, the structural aspect identifies the connections and relationships between those functional components.

Early theories, such as the "method of systemic design" [53], were based on a linear structure with three major functions: analysis, synthesis, and evaluation. In general, there is a strong line from the systematic design method that inspires our DSR block as a linear, step-based process. However, the DSR block additionally has several critical

[53]: Jones (1963), "A method of systemic design"

differences that result from the unique challenges of crowdsourcing and the inclusion of non-experts in the process. First, the design block encapsulates a single designer's work process that includes a personal design sub-process. This sub-process is personal and is not in the scope of crowdsourcing research [11]. Second, since evaluation is not only the selection of the best solutions, but also a reflective feedback process that helps to improve the selected artifacts, the evaluation step in our DSR block is divided into selection and review micro-tasks.

[11]: Maher (2011), "Design Creativity Research: From the Individual to the Crowd"

According to Luckman (1967), due to the interdependent nature of the design problem elements, the design process is cyclical rather than linear [54]. This results in an agreed two-dimensional model where the systemic design method is embedded as a component of a larger design process. This concept is similar to our crowdsourcing model in terms of having two design process hierarchies and the defined design stage—namely, preliminary, sketch, and detailed design. The later process influences RIBA's "plan of work", a popular architectural design process that includes concept, developed, and technical design stages. However, in the present study, we found out that the two-dimensional model falls short in micro-task-based crowdsourcing, as three of the resulting artifacts were still preliminary and required further development. Our results revealed that the design process requires multiple iterative [56]. However, we relied on human decisions to signal the completion of the design process.

[54]: Luckman (1967), "An Approach to the Management of Design"

[56]: Kline (1985), "Innovation Is Not a Linear Process"

According to Maher's (1996) co-evolution computational model, the problem-space evolves together with the solution space. We believe that the DSR block follows this notion with the review micro-task [8]. Specifically, by generating reviews, the process participants actively add information to the design project, thereby co-evolving the problem space.

[8]: Maher et al. (1996), "Formalising Design Exploration as Coevolution: A Combined Gene Approach"

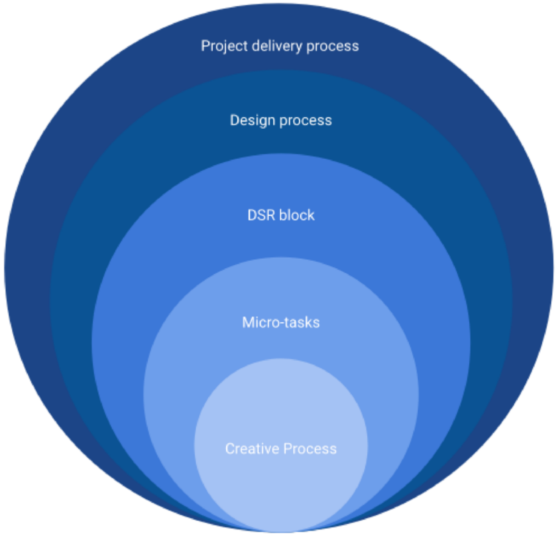


Figure 6.12: Overview of the Design Process

6.3.1 Contribution to Design theory

Design is a process that involves cognitive and social interaction and therefore is complex. While the mentioned design process models help understand specific aspects of the design process, they are inadequate as an algorithmic design process. Such a model was our first hypothesis, that did not perform well. Therefore, we contribute a design process model that is suitable for algorithmic and open-sources processes and thus more complex and handles the realistic challenges of a design process.

The models mentioned in the literature review described an iterative design process (see Section 2.1). Additionally, there is a linear project delivery process overarching the previous process (see Section 2.2). Our process adds three layers between these models that result in a more complex process that consists of five layers (See Figure 6.12): 1) Cognitive Process, 2) Micro-Tasks, 3) DSR Block, 4) Design Process, and 5) Project Delivery Process.

The model starts with a human being at the core of the process, going through a *cognitive process* in their mind, which is crucial to the process's success. Each participant experiences the process differently, influenced by various variables, particularly information and knowledge. Simple *micro-tasks* encapsulate the cognitive processes. Each micro-task is defined by input information, task process, and expected output, similar to a computer function. Next, the *DSR Blocks* encapsulate several dozen micro-tasks in a design exploration process. First, several design micro-tasks explore the solution possibilities. Next, multiple selection micro-tasks filter the best solutions. Finally, review micro-tasks add essential information that updates the problem space and provides directions for further exploration. On the next process layer, the *design process* iterates between various architectural media and consists of several DSR blocks. Finally, the linear *project delivery process* organizes the phases of the project toward finalizing the design. In conclusion, our model contributes a more complex view of the design process, which is more applicable as a crowdsourcing algorithm.

6.4 Quality Considerations

Quality is a complex and open issue in architecture. Since any study on crowdsourcing requires a method to choose the best fitting artifact, in what follows, we discuss quality considerations in creative crowdsourcing from the HCI perspective.

Quality of design stems above all from the individual abilities of designers. In crowdsourcing, we combine these abilities to generate a whole that is larger than its parts. To achieve an artifact that is a product of genuine collaboration, we provide several considerations identified in the present study.

6.4.1 Multiplicity Considerations

In the present study, quality was achieved through the multiplicity of design tasks and artifacts. The selection process highlighted the best designs. As discussed previously, this approach is similar to the conventional three-stage systemic design process [53]. However, since crowdsourcing is an internet method, it can connect more people, from different cultures, without discrimination and prejudice, all of which makes exploration more diverse and solutions more creative. Although the process of selecting the best fitting artifacts remains a challenge, in the present study, we assumed that a selection process would provide the best results. However, it was also deemed reasonable that more than one artifact could lead to a good design, which suggests that it would be wasteful to dismiss good designs at the preliminary design stages since those designs could also be the basis for the final artifact. Given the vast human potential in crowdsourcing, we conclude that multiplicity in the design process will foster design quality.

[53]: Jones (1963), "A method of systemic design"

Figure 6.13 shows the artifacts produced in Experiments 16-25 analyzed as a hierarchical tree with nine levels. Each node in Figure 6.7 corresponds to one artifact and has a number that identifies its designer. While some designs are further explored, most are not and have no leaves. It can be seen that the artifacts of the final design are based on multiple artifacts produced by several designers.

Given there is a group of designers, some of whom have superior design skills, we can reasonably predict that the final output would be made of the contributions of these individual designers. However, in the experiments with a diverse group of architects and students, the results demonstrated that the final artifacts were joint products of seven different designers.

6.4.2 Individual Consideration

The results of Experiment 1 demonstrated that the participants found it stressful to receive a brief right before the sketching task, which eventually led to the production of low-quality artifacts. However, when the brief was provided in advance (i.e., a day earlier), the participants showed a better performance and were more effectively

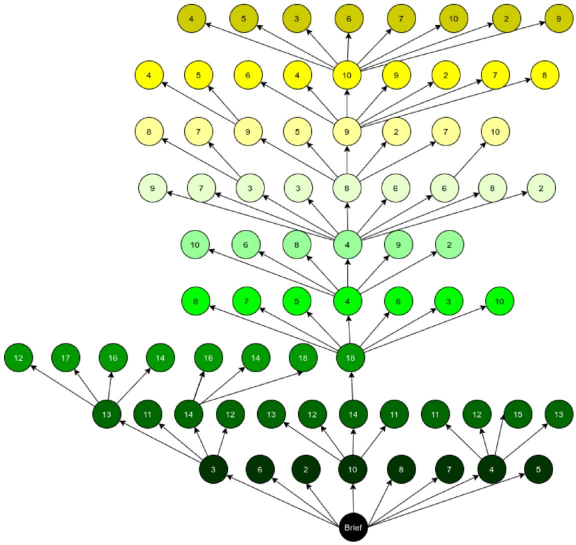


Figure 6.13: Experiment 16-25 artifact tree. Each circle represents a number that identifies a specific designer. The artifacts are ordered in rows representing design improvement iterations.

focused on the design. From these results, we can conclude that, in design crowdsourcing tasks, additional time should be provided to the designers to "sleep over it" before they are asked to start working on the design.

Overall, the transitions between different types of architectural artifacts in the design process led the participants to demonstrate more creativity. For instance, a conceptual sketch can be developed as various plans. The most significant evidence for this was obtained in Experiment 25 that aimed to improve the designs by the transition from a CAD plan to a sketch.

As mentioned above, the participants enjoyed using the software, creating designs and were motivated by the spirit of competition. Based on these results, we can conclude that competition promoted the achievement of better results (i.e., higher-quality artifacts).

6.5 Participatory Design Process

In this section, we address community and stakeholder participation in the process. Our initial goal was to develop a new collaborative design method based on open-source theories in architecture [40]. To this end, the following research question [RQ-5] was formulated: *"In which parts of the architectural crowdsourcing workflow do professional participants provide better performance and results than non-professional participants?"*

[40]: Dortheimer et al. (2020), "Open-source architecture and questions of intellectual property, tacit knowledge, and liability"

Accordingly, the experiments were designed to identify which tasks would help integrate designers and stakeholders into a design process and to experimentally determine which tasks non-expert stakeholders and community members can participate in.

As mentioned above, the results revealed that design micro-tasks should be performed by architects since the architectural design is a skill that takes considerable time to acquire and hone. Architects also know how to meet complex requirements, generate quality sketches, and operate sophisticated CAD software.

Since our study participants were architecture students and professionals, so neither of these two groups could be referred to as laypeople (i.e., non-experts). However, as revealed by the relationship between experience and quality of deliverables, some participants — particularly those without extensive experience — underperformed on the design tasks. From this evidence, we can conclude that the participants with no relevant experience can be expected to perform poorly.

However, since selecting specific designs does not necessarily require design skills, non-expert stakeholders may also successfully perform the task of selecting of most suitable artifacts. This conclusion is based on the results of Experiment 9 on the correlation between non-expert and expert evaluations. While professional architects may have a broader understanding of the quality of a design, in the present study, we reasoned that, due to a better knowledge of the project's environment, culture, and context, non-expert stakeholders' input would be as significant as that of experts.

Similarly to selection micro-tasks, review micro-tasks can also be performed by both experts and non-experts. However, reviews by non-expert participants should be complemented with experts' reviews, as non-professionals may not be aware of all the shortcomings of the design.

As mentioned previously, many previous studies argued the disconnection between architects and end-users is a major cause of why large-scale projects fail (see Section 2.2.2). However, although the engagement of non-experts in complex design and engineering projects remains a challenge [75], crowdsourcing offers new potential in tackling this challenge by collecting stakeholders' opinions and ideas [108]. Our experimental results provide substantial evidence that "the wisdom of the crowd" largely depends on the specific method applied in a given micro-task (e.g., selection vs. rating vs. comparison). Using a selection method can produce a more explicit crowd assessment that would strongly correlate with expert evaluation.

[75]: Robertson et al. (2012), "Challenges and Opportunities in Contemporary Participatory Design"

[108]: Brabham (2009), "Crowdsourcing the Public Participation Process for Planning Projects"

6.6 Architecture Competitions and Practice

In this section, we discuss the implications of crowdsourcing technology on architectural practice. As argued in Section 1.1, since architecture is a knowledge-based practice, it can be improved and revolutionized using information technologies like crowdsourcing and AI.

Continuing architectural tradition and previous crowdsourcing studies, the proposed process includes micro design competitions. While architectural competitions usually have one or two stages, the crowdsourcing process proposed and applied in the present study included multiple iterations that allowed for an agile, dynamic, and more collaborative design process.

Since competitions do not offer fair compensation for participants' efforts, they can be regarded to be exploitative

[49]: Deamer (2015), "The Guggenheim Helsinki Competition: What Is the Value Proposition?"
[50]: Keslacy (2018), "Arcbazar and the Ethics of Crowdsourcing Architecture"
[109]: Ekbia et al. (2017), *Heteromation, and Other Stories of Computing and Capitalism*

[49] and thus problematic in terms of their application on crowdsourcing websites [50]. Using competitions, capitalism can take advantage of a cheap and desperate workforce under the false promise of international publicity and success [109]. The research on online technologies to produce labor is still growing, and these sites are the first generation of crowdsourcing sites based on outdated technology.

In contrast to the above, the present study reveals the potential of crowdsourcing as a fair way to online and collaborative work in architecture. The contest that underlies the crowdsourcing process has no monetary prize, as all participants are paid for their efforts. Moreover, a collaborative creative dynamic is formed. As demonstrated by our experimental results, in most cases, the work of different designers was selected. Designers evaluate other designers' work and create derivative works. Gradually, different people make their creative contributions. In doing so, the design process expands beyond the crowdsourcing process embedded in the software and can be seen as a more general framework for any organized design process.

Furthermore, from our casual discussions with the participants, we learned that they experienced high satisfaction in the creation and enjoyed the spot-like competition. This allows us to conclude that the presented model may facilitate innovation in architecture, fair distribution of rewards among contributors, minimizing the risks, and addressing ethical issues typical of architectural competitions [49] and commercial crowdsourcing websites [50].

[49]: Deamer (2015), "The Guggenheim Helsinki Competition: What Is the Value Proposition?"
[50]: Keslacy (2018), "Arcbazar and the Ethics of Crowdsourcing Architecture"

Based on the findings, we argued that the methods used in the present study could be meaningfully applied in several ways. First, small architectural firms can harness a crowd of architects in the initial design stages to generate creative ideas. Second, at more advanced stages of the design process, architects can be assisted by the crowd to solve various parts of the project in further detail.

This approach can be seen as an organization in the "cloud" — that is, a virtual team that can grow and shrink depending on the firm's business constraints. In addition, this approach can also help early-stage architects compete on larger projects and compete with established firms.

Furthermore, the proposed approach can be used to manage creative processes in medium and large architectural firms. Since the process is anonymous, office workers can share ideas and choose the most promising ideas without pressure from the organizational power structure. It would also be possible to enable a hybrid operation mode where office workers and crowd-workers transparently collaborate on a joint project.

Definitely, our approach is valuable in terms of engaging project stakeholders, clients, various professionals, neighbors, and the community at large to benefit from the tacit knowledge and contribute to the architectural outcome. Our method is simple and accessible to non-experts and can help to find a new and desirable form of integrative and meaningful public participation.

However, crowdsourcing alone cannot replace an architect who knows the client, plot, relevant regulations, target community, language, and local culture. Ultimately, it is the responsibility of the local architect to build a relationship with clients and win their trust in the most significant and expensive projects in their life. Therefore, our point is, rather than as an end-user-facing technology, crowdsourcing should serve as an organizational technology of architectural firms.

In conclusion, in the present study, we have proposed several different ways of using crowdsourcing in architectural design. Our results convincingly demonstrated that crowdsourcing in architecture has a strong potential for fairer and more efficient use and healthier competition. This having been said, we conclude by reinstating that, given that technologies do not include a moral system and are a tool in a complex political system, the ultimate responsibility should explicitly lie with the humans who use it.

6.7 Open Source Architecture

The open-source movement is a concept, a philosophy, and a political movement aimed at disrupting intellectual property laws. Benefiting from this approach, thousands of programmers have formed communities and developed software that has become the infrastructure and the foundation of the Internet and computing today. Furthermore, the open-source programmers have built novel tools and methods that enable geographically distributed and asynchronous software development and management to support their projects.

On the verge of the second millennium, the open-source movement captured the imagination of Silicon Valley and quickly spread to other fields, including, among many others, architecture and urban planning. Soon after the first appearance of an open-source architecture text in 2003, various authors made significant contributions to the topic in subsequent years. In his speech upon receiving the Pritzker Prize, Alejandro Aravena said that he was releasing architectural plans as an open source for the benefit of humanity. However, although open-source architecture has been debated over the past 18 years, its definition remains unclear. Is open-source providing architectural plans for free, as suggested by Alejandro Aravena? Or is the new role of the architect to lead clients in chorus [36]? Or yet, is this a new ownership model for the city [110]?

[36]: Ratti et al. (2015), *Open Source Architecture*

[110]: Fuller et al. (2008), *Urban Versioning System 1.0*

In a study on open-source architectural projects, all projects that called themselves "open-source" were found to lack open-source characteristics [40]. One specific failure was that there was no significant user community, suggesting that the development of a creative architectural community was hindered by the lack of tools and methods to collectively create architecture. However, one such method is crowdsourcing, a distributed production method that can collaboratively create architecture using information technologies.

[40]: Dortheimer et al. (2020), "Open-source architecture and questions of intellectual property, tacit knowledge, and liability"

6.7.1 Architectural crowdsourcing as a method of open-source production

The essay "The Cathedral and Bazaar" by Eric Raymond is a cornerstone of the open-source movement [28]. In this essay, Raymond (1999) identifies organizational insights from the programmer communities that allowed for the development of complex software in a novel way.

[28]: Raymond (1999), *The cathedral and the bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*

Crowdsourcing supports many of open-source principles discussed in Section 1 of Raymond's (1999) essay. By integrating stakeholders and clients in selection and review micro-tasks, crowdsourcing enables one to treat various users as co-developers. The clients—provided they possess necessary skills—can even participate in the design process. Overall, since crowdsourcing is based on significant information from the customers, this approach ensures that customers' voices are heard and incorporated.

Furthermore, Raymond (1999) suggests having a sufficiently large user and co-developer base to allow access to many people and increase the chance of solving problems. Similarly, he also argues that, in the long run, many heads are better than one. Although we have not tested how the system would work with thousands of users, there is evidence that it works well with dozens and may scale to handle more users. Dealing with dozens of clients is not easy in architectural design processes. However, since our method allows many people to participate in the process, principally through the review micro-task, it offers a valuable opportunity to explore more solutions to design problems.

Finally, Raymond (1999) argues the need to identify good ideas from users. Accordingly, using our proposed process, the crowd can be used to identify good design ideas for the selection micro-tasks by using the wisdom of the crowd. While this process emphasizes majority decision-making, it also permits the simultaneous development of multiple ideas [2]. In this way, good ideas can be developed and be re-evaluated in the next DSR block.

[2]: Carpo (2011), *The Alphabet and the Algorithm*

Along with Raymond's (1999) insights regarding the benefits afforded by using clients' and stakeholders' knowledge, the crowdsourcing process can also promote democratic decision-making, which offers a sense of participation and agreement.

Of note, however, while some "open source" approaches in architecture claim to seek to bridge the knowledge gap between clients and architects [36], this has nothing to do with open source communities. The software is written by skilled programmers, not by the clients. Instead, clients contribute knowledge through discussions, including suggestions for new features, improvements, and bug reporting. In this sense, the crowdsourcing process is similar to software open source communities.

[36]: Ratti et al. (2015), *Open Source Architecture*

6.7.2 Redefining open-source architecture

Based on the results of this study, we propose the notion of open source in architecture — a novel notion that conceptualizes architecture as a knowledge-rich profession. In *open-source architecture*, information technologies are leveraged to empower designers' and users' active participation in shared and collaborative human-centered design processes, with the ultimate outcome of fostering collective intelligence. Accordingly, the architect is no longer framed as a hero or artist whose artistic expression is at its center. Instead, open-source architecture is best described as architecture that emerges within a community and, through the use of advanced information and communication technologies in the search for good design, makes full use of the cognitive and creative abilities of relevant stakeholders, including architects and clients.

6.8 Limitations

The present study has several limitations. First, since our primary focus was on disassembling the design process, we were not concerned with decomposing the architectural design problem, solving it, and merging the solutions using crowdsourcing. In essence, the decomposition of the design problem is a central topic in the computation of design. While there is substantial literature in the field of design and AI in the 20th century, relevant publications on crowdsourcing are scarce [111]. Therefore, in future research, it would be meaningful to explore crowdsourcing methods in solving such problems.

[111]: Kulkarni et al. (2011), "Turkomatic : Automatic Recursive Task and Workflow Design for Mechanical Turk"

Second, the model proposed in the present study was developed under "laboratory conditions" with students and architects. Due to budget constraints, we could not test the model and software under the conditions of a realistic project through empirical experiments. Accordingly, further research is needed that would examine the model under the terms of a real architectural design project. In such a project, professional architects and a process manager (e.g., a developer or chief architect) would use the software. Such a study could provide new insights that may enhance the model.

Despite the limitations mentioned above, as well as the difficulties encountered during the experiments—such as the challenges of bias, design thinking, graphic user interface, technical challenges, project management, the scale of data, perception, and so on—the model proposed in the present study is firmly established, as it was tested in four different projects and multiple experiments. The conditions under which the present model was developed created many realistic challenges in crowdsourcing. For instance, since most participants were students with limited experience, the quality of their design products varied as compared to that of professional architects, highlighting the challenges in selecting the best artifacts and the knowledge gap with unprofessional designs.

7 Conclusion

Architecture is an ancient profession that has undergone many changes throughout history. Owing to the current industrial revolution focusing on information technologies and the fact that architecture is information, there is tremendous untapped potential for an open-source architecture.

One such potential is crowdsourcing technology that offers mechanisms to capture collective intelligence. This technology may offer necessary improvements to architectural design, such as a more participatory, democratic, high-quality, efficient, and economical design process. Therefore, the dissertation aimed to develop a new architectural design method based on concepts and technologies from the field of crowdsourcing.

We introduced a new architectural design method based on breaking up the architect's design process into micro-tasks. A crowd of participants performs the micro-tasks. Finally, the output is merged into a coherent work of architecture. We have shown that architectural artifacts can be created with this method and formulated a theoretic three-level model. The presented design process was based on a transition from sketches and computer artifacts, thus encouraging creativity and design thinking.

We explored the crowdsourcing process and determined how architects and stakeholders can collaborate through the design process. The proposed approach considers the tacit knowledge of diverse participants.

We presented the development of simple selection and review tasks, all of which resulted in a clear collective selection of the most suitable artifacts. Through these tasks, the crowdsourcing process can become more appropriate for participatory design and enhance stakeholders' engagement.

The design process was based on an idea tree that facilitated the development of a diversity of design solutions simultaneously while displaying dynamics that narrowed the idea branches through stakeholders and community participation.

The DSR block model was introduced as a framework for design crowdsourcing processes. This block supports searching for solutions, followed by an effective and efficient reduction of those solutions, and finally, collecting reviews from the participants.

Furthermore, we developed an architectural design microtask based on a compact project brief with the critical information needed to participate in the design process. Architectural design tasks also of output examples, demonstration videos, and simple task steps.

Taken together, the results of the present study contribute to previous research on design crowdsourcing, collaborative design, participatory design processes, and open-source architecture. The results show high potential in implementing crowdsourcing methods in architectural design that harness the collective intelligence of architects and designers and the stakeholders' in-depth knowledge of the 'genius loci.'

Appendix

Application Screens | A

A.1 Version 1.0

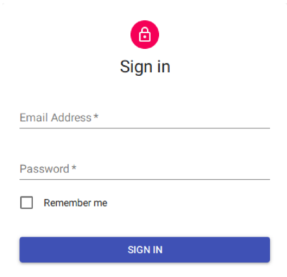


Figure A.1: Log-in screen

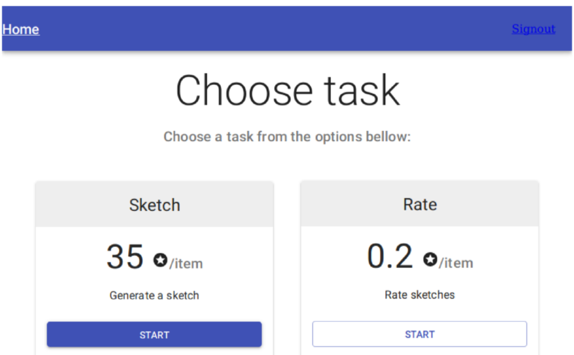


Figure A.2: Select task screen

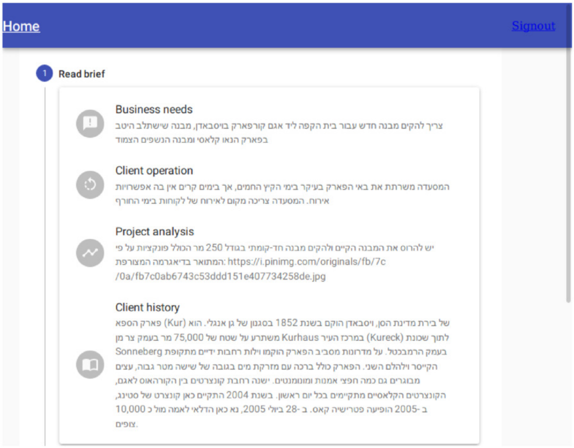


Figure A.3: Task steps screen: Brief

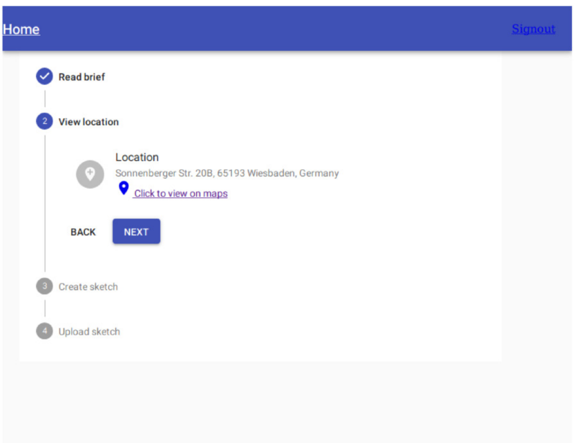


Figure A.4: Task steps screen: Location

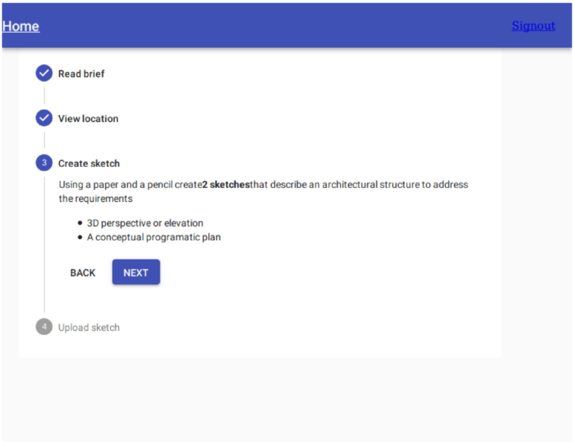


Figure A.5: Task steps screen: Task

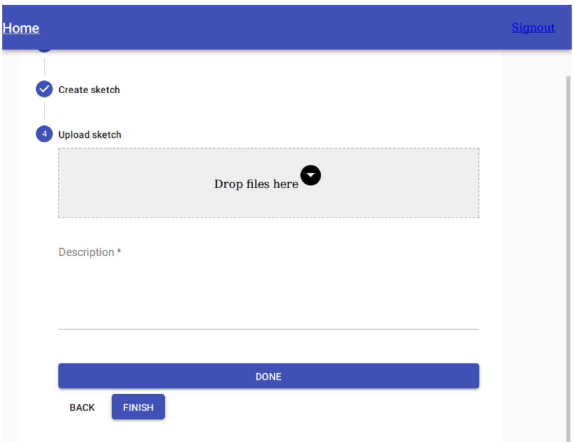


Figure A.6: Task steps screen: Upload

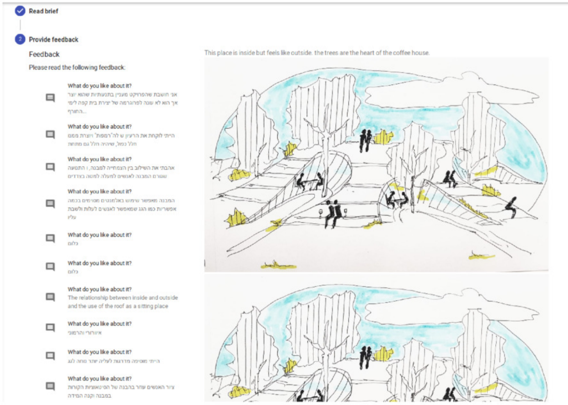


Figure A.7: Provide review screen

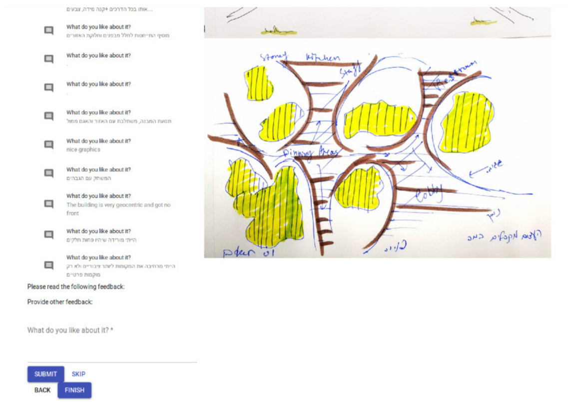


Figure A.8: Provide review screen - continued

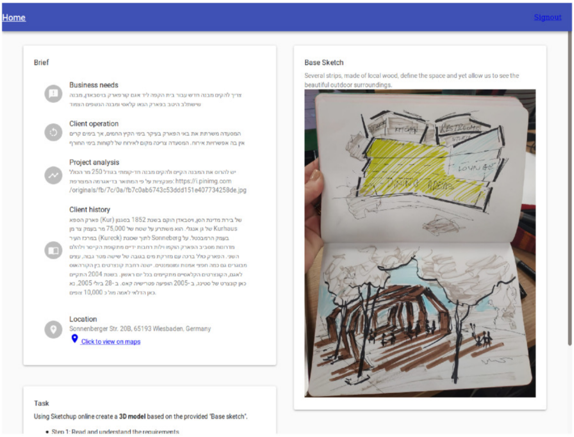


Figure A.9: Task steps screen: create model

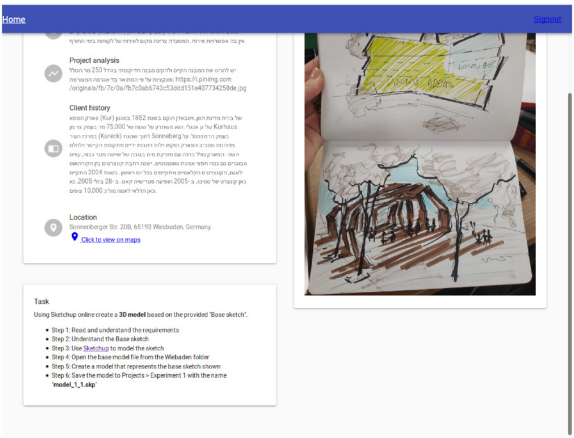


Figure A.10: Task steps screen: create model - continued

A.2 Version 1.1

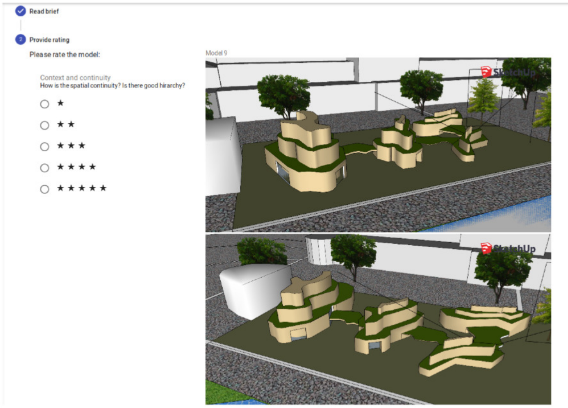


Figure A.11: Selection task screen

A.3 Version 1.2

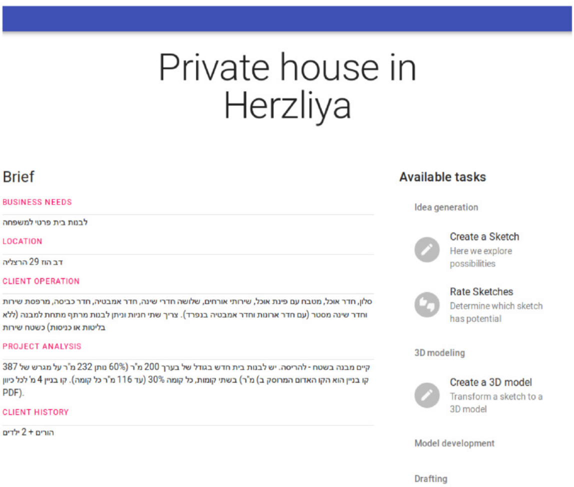
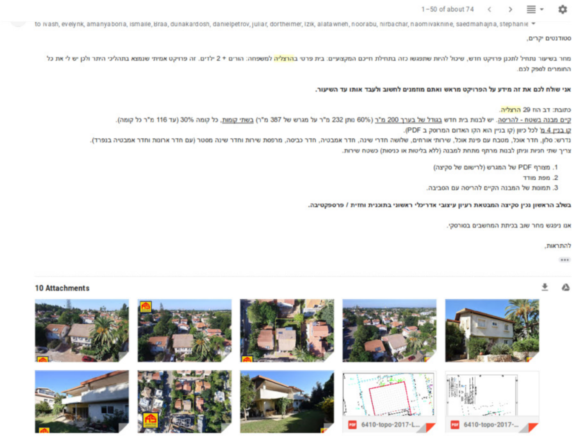


Figure A.12: Brief screen and task selection



1–50 of about 74

to ivash, evelynk, amanyaboria, ismaile, Braa, dunakardosh, danielpetrov, juliar, dortheimer, izik, alatawneh, noorabu, nirbachar, naomivaknine, saedmahajna, stephanie

סטודנטים יקרים,

מחר בשיעור נתחיל לתכנן פרויקט חדש, שיכול להיות שתפגשו כזה בתחילת חייכם המקצועיים: בית פרטי בהרצליה למשפחה: הורים + 2 ילדים. זה פרויקט אמיתי שנמצא בתהליכי היתר ולכן יש לי את כל החומרים לספק לכם.

אני שולח לכם את זה מידע על הפרויקט מראש ואתם מוזמנים לחשוב ולעבד אותו עד השיעור.

כתובת: דב הוז 29 הרצליה.
קיים מבנה בשטח - להריסה. יש לבנות בית חדש בגודל של בערך 200 מ"ר (60% נותן 232 מ"ר על מגרש של 387 מ"ר) בשתי קומות, כל קומה 30% (עד 116 מ"ר כל קומה).
קו בניין 4 מ' לכל כיוון (קו בניין הוא הקו האדום המרוסק ב PDF).
נדרש: סלון, חדר אוכל, מטבח עם פינת אוכל, שירותי אורחים, שלושה חדרי שינה, חדר אמבטיה, חדר כביסה, מרפסת שירות וחדר שינה מסטר (עם חדר ארונות וחדר אמבטיה בנפרד).
צריך שתי חניות וניתן לבנות מרתף מתחת למבנה (ללא בליטות או כניסות) כשטח שירות.

1. מצורף PDF של המגרש (לרישום של סקיצה)
2. מפת מודד
3. תמונות של המבנה הקיים להריסה עם הסביבה.

בשלב הראשון נכין סקיצה המבטאת רעיון עיצובי אדריכלי ראשוני בתוכנית וחזית / פרספקטיבה.

אנו ניפגש מחר שוב בכיתת המחשבים בסורסקי.

להתראות,

10 Attachments

Figure A.13: Brief email

Figure A.14: Sketch task steps

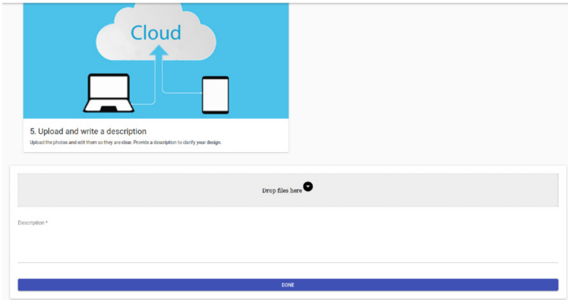


Figure A.15: Sketch task upload

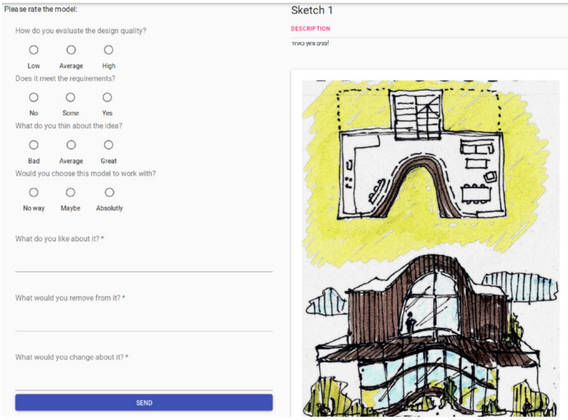


Figure A.16: Rate and review task screen

A.4 Version 1.3

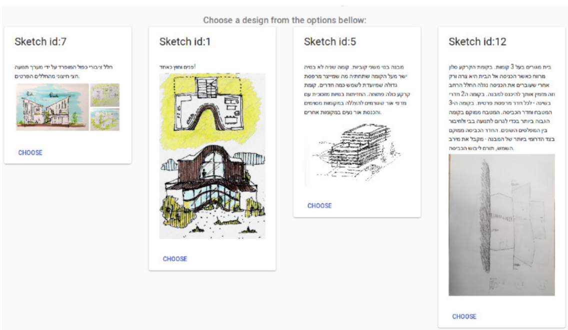


Figure A.17: Selection base artifact screen

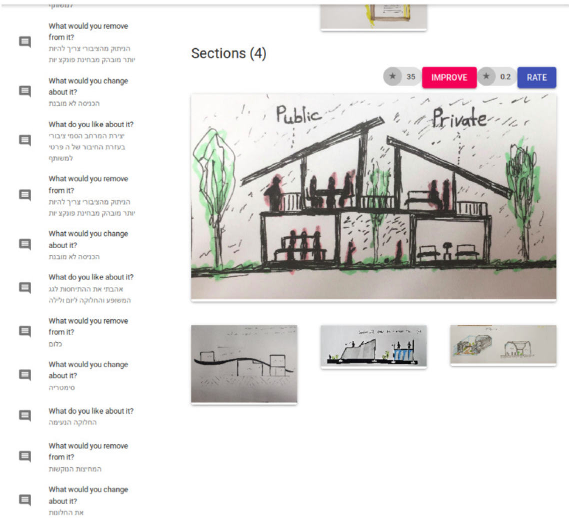


Figure A.18: View base artifacts and critique for merge screen

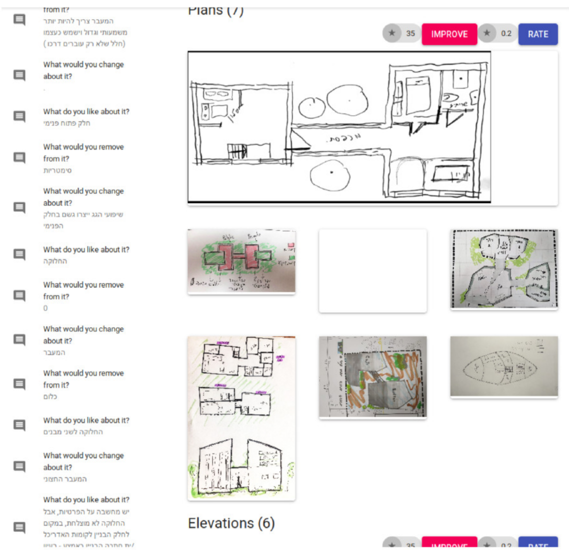


Figure A.19: View base artifacts and critique for merge screen

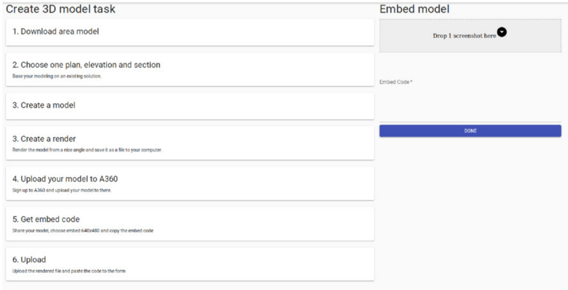


Figure A.20: Task steps screen

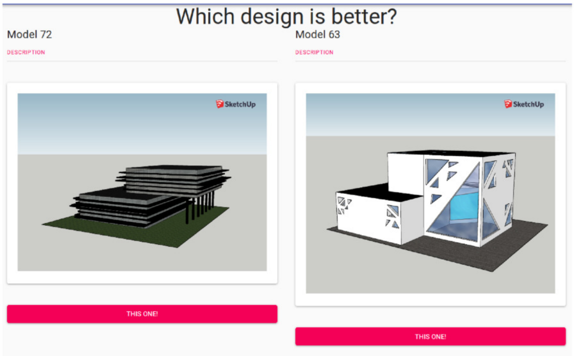


Figure A.21: Selection task screen

A.5 Version 1.4

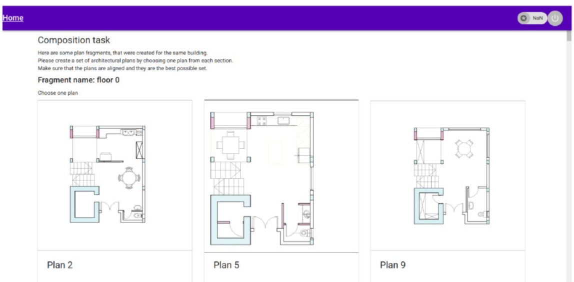


Figure A.22: Composition task screen

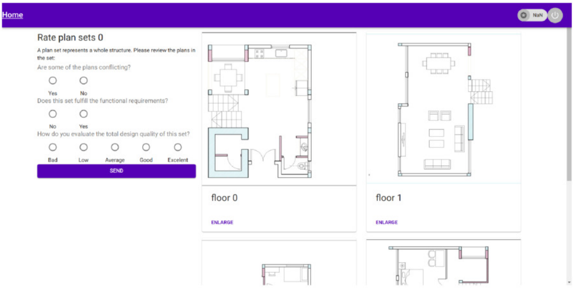


Figure A.23: Set rating task screen

A.6 Version 2.0

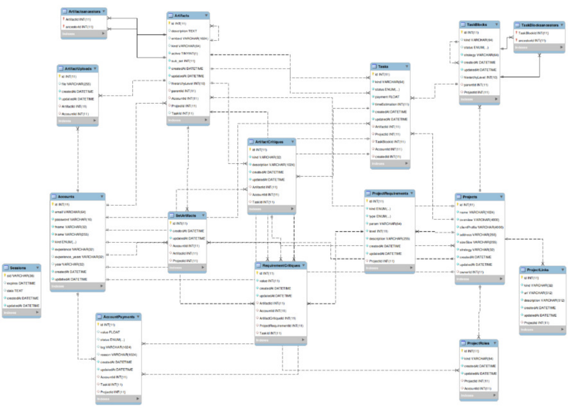


Figure A.24: Database model diagram

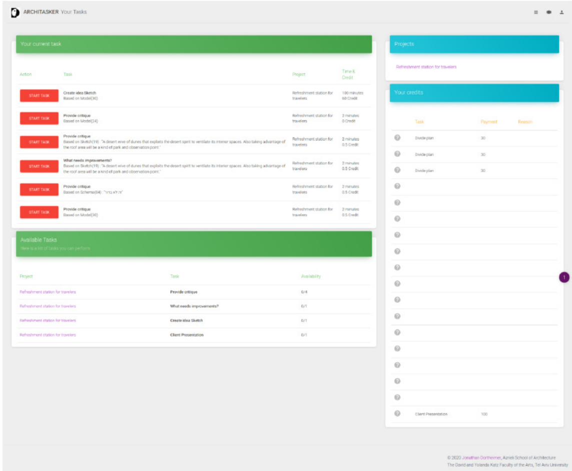


Figure A.25: Main user dashboard where tasks can be viewed

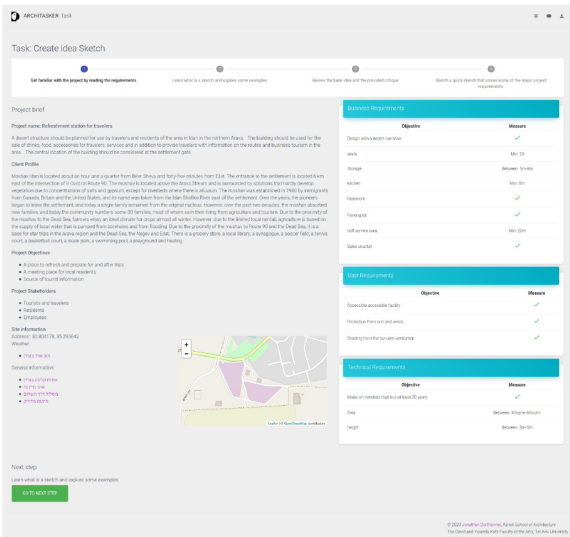


Figure A.26: Task brief screen

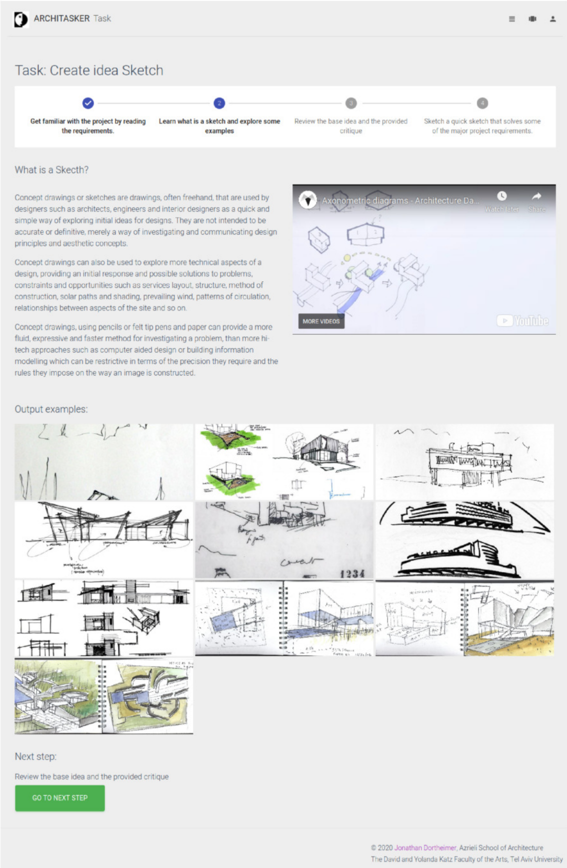


Figure A.27: Task example output screen

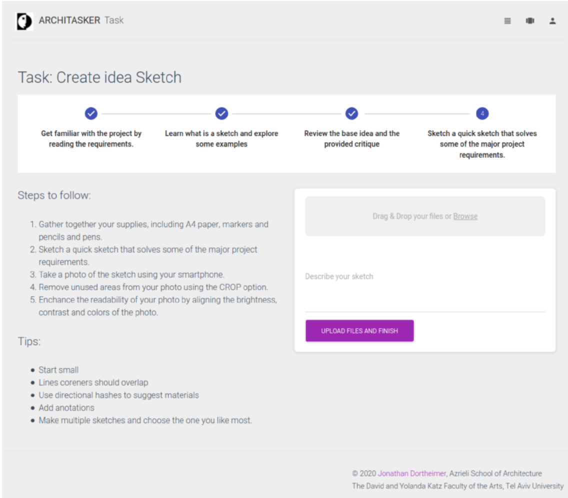


Figure A.28: Task upload artifact screen

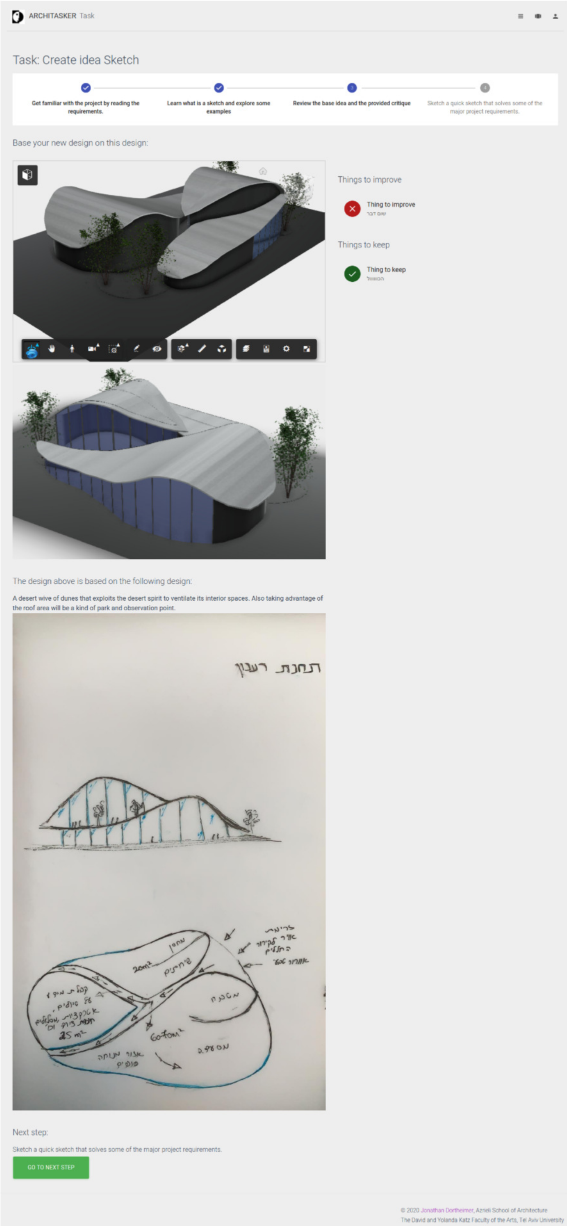


Figure A.29: Task artifact to improve screen

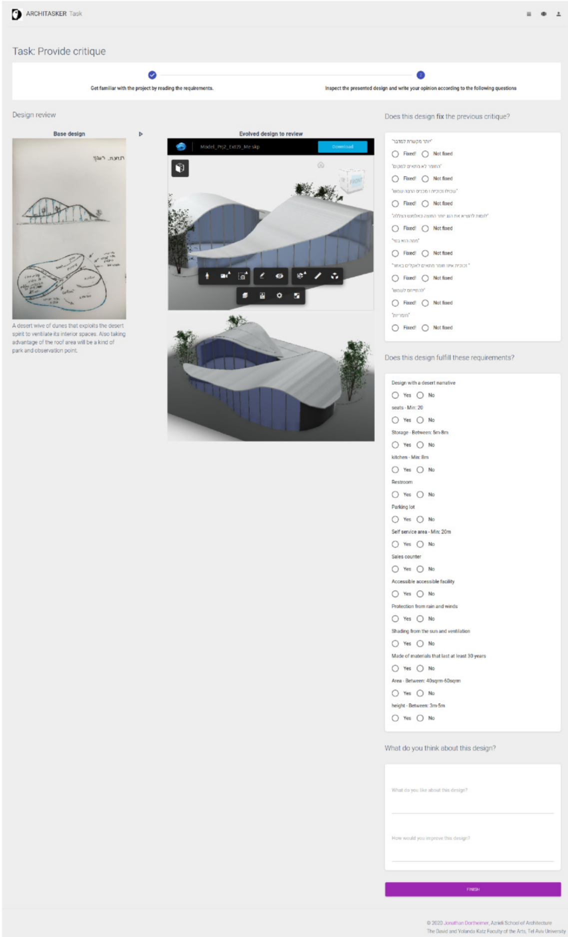


Figure A.30: Review task screen

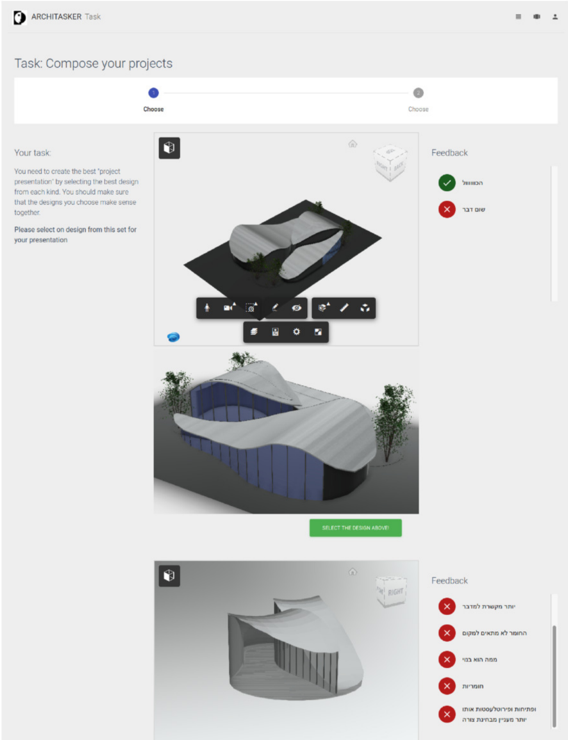


Figure A.31: Selection task screen

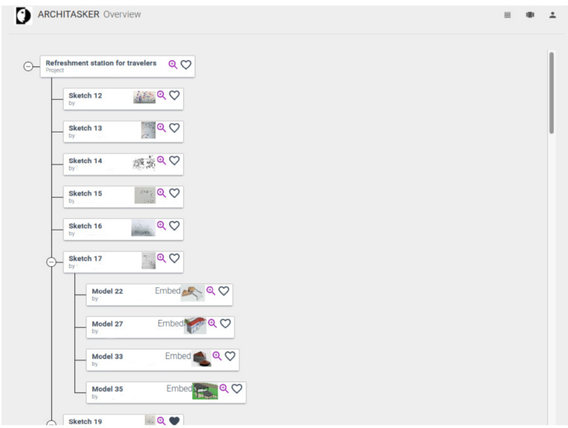


Figure A.32: Artifact design tree screen

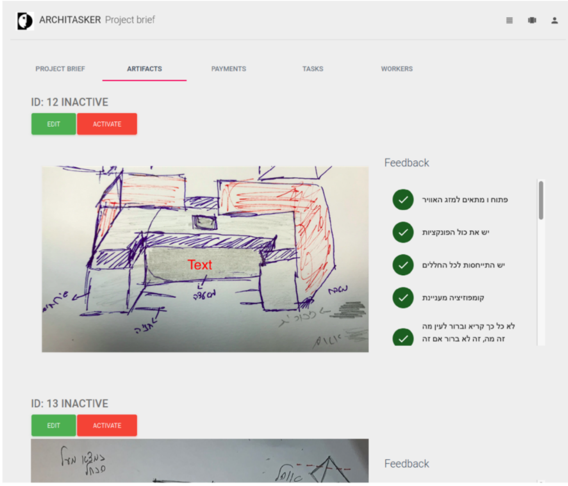


Figure A.33: Artifacts administration screen

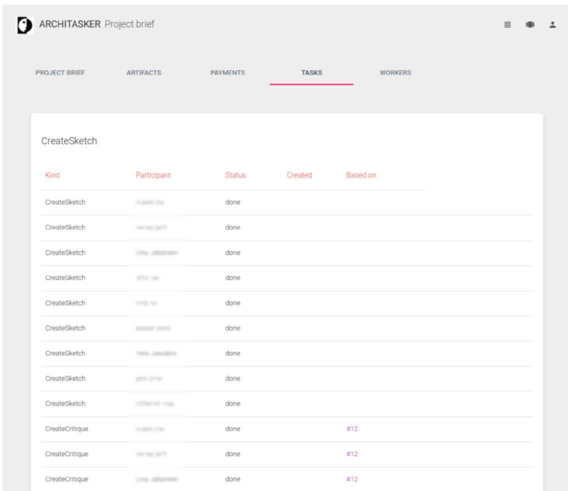


Figure A.34: Task administration screen

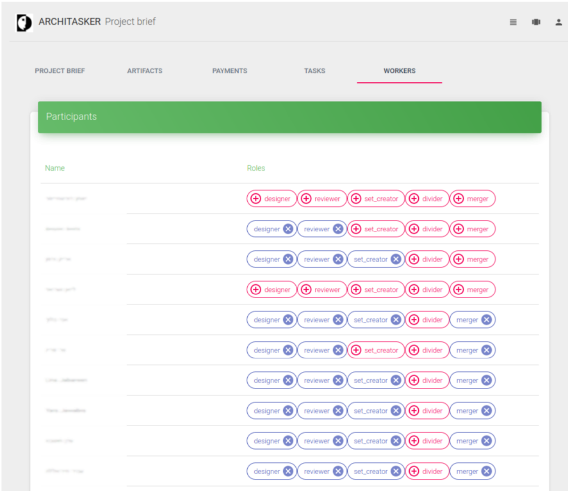


Figure A.35: Participant administration screen (Names are blurred for privacy)

Bibliography

Here are the references in citation order.

[1] Pollio Marcus Vitruvius. *The Architecture of Marcus Vitruvius Pollio*. London: Lockwood & co., 1874 (cited on pages 2, 28).

[2] Mario Carpo. *The Alphabet and the Algorithm*. MIT Press, 2011, p. 184 (cited on pages 2, 167, 211, 225).

[3] Marc Uri Porat. *The Information Economy: Definition and Measurement*. Tech. rep. Office of Telecommunications (DOC) , Washington, D.C., 1977, p. 319 (cited on page 2).

[4] Erich Gamma et al. *Design Patterns: Elements of Reusable Object-Oriented Software*. 1st ed. Addison-Wesley Professional, 1994, p. 416 (cited on page 3).

[5] Herbert A. Simon. *The Sciences of the Artificial*. MIT Press, 1969, p. 123 (cited on pages 3, 23, 54, 60, 214).

[6] Herbert A. Simon. "The structure of ill structured problems". In: *Artificial Intelligence* 4.3-4 (1973), pp. 181–201. DOI: 10.1016/0004-3702(73)90011-8 (cited on page 3).

[7] Hideaki Takeda et al. "Modeling design processes". In: *AI Magazine* 11.4 (1990), pp. 37–48. DOI: 10.1609/aimag.v11i4.855 (cited on pages 3, 23).

[8] Mary Lou Maher, Josiah Poon, and Sylvie Boulanger. "Formalising Design Exploration as Co-evolution: A Combined Gene Approach". In: *Advances in Formal Design Methods for CAD: Proceedings of the IFIP WG5.2 Workshop on Formal Design Methods for Computer-Aided Design* (1996), pp. 3–30. DOI: 10.1007/978-0-387-34925-1_1 (cited on pages 3, 24, 215).

[9] Kees Dorst and Nigel Cross. "Creativity in the design process: Co-evolution of problem-solution". In: *Design Studies* 22.5 (2001), pp. 425–437. DOI: 10.1016/S0142-694X(01)00009-6 (cited on pages 3, 24).

[10] John S. Gero and Udo Kannengiesser. "The situated function-behaviour-structure framework". In: *Design Studies* 25.4 (2004), pp. 373–391. DOI: 10.1016/j.destud.2003.10.010 (cited on page 3).

[11] Mary L. Maher. "Design Creativity Research: From the Individual to the Crowd". In: *Design Creativity 2010*. London: Springer, 2011, pp. 41–47. DOI: 10.1007/978-0-85729-224-7_7 (cited on pages 4, 25, 215).

[12] Antoine Picon. "From Authorship to Ownership: A Historical Perspective". In: *Architectural Design* 86.5 (2016), pp. 36–41. DOI: 10.1002/ad.2086 (cited on page 4).

[13] Imdat As, Siddharth Pal, and Prithwish Basu. "Artificial intelligence in architecture: Generating conceptual design via deep learning". In: *International Journal of Architectural Computing* 16.4 (2018), pp. 306–327. DOI: 10.1177/1478077118800982 (cited on page 4).

[14] Stanislas Chaillou. "AI + Architecture | Towards a New Approach". PhD thesis. Harvard University, 2019. DOI: 10.9783/9781949057027-006 (cited on page 4).

[15] Janet McDonnell. "Collaborative negotiation in design: A study of design conversations between architect and building users". In: *CoDesign* 5.1 (2009), pp. 35–50. DOI: 10.1080/15710880802492862 (cited on page 5).

[16] Antoine Picon. "From Authorship to Ownership: A Historical Perspective". In: *Architectural Design* 86.5 (2016), pp. 36–41. DOI: 10.1002/ad.2086 (cited on pages 5, 11).

[17] Ryan E. Smith, Erin Carraher, and Peter DeLisle. *Leading Collaborative Architectural Practice*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2017 (cited on pages 5, 29).

[18] Antoine Picon. *Ornament: The Politics of Architecture and Subjectivity*. John Wiley & Sons Ltd, 2013 (cited on page 5).

[19] Andrew Saint. *The Image Of The Architect*. 1st. Yale University Press, 1983, p. 180 (cited on page 5).

[20] Bruno Vayssiere and Christopher Alexander. "Notes on the Synthesis of Form". In: *Leonardo* 10.3 (1977), p. 257. DOI: 10.2307/1573476 (cited on pages 6, 28, 29).

[21] Jane Jacobs. *Death and life of great American cities*. New York: Vintage books, 1961, p. 458 (cited on pages 6, 30).

[22] Bernard Rudofsky. "Architecture without Architects: A Short Introduction to Non-Pedigreed Architecture". In: *Architecture without Architects*. New York: The Museum of Modern Art, 1965, p. 71 (cited on page 6).

[23] Yoram Reich et al. "Varieties and issues of participation and design". In: *Design Studies* 17.2 (1996), pp. 165–180. DOI: 10.1016/0142-694X(95)00000-H (cited on page 6).

[24] Elizabeth B.-N. Sanders and Pieter Jan Stappers. "Co-creation and the new landscapes of design". In: *CoDesign* 4.1 (2008), pp. 5–18. DOI: 10.1080/15710880701875068 (cited on pages 6, 29).

[25] Elizabeth B.-N. Sanders, Eva Brandt, and Thomas Binder. "A framework for organizing the tools and techniques of Participatory Design". In: *ACM International Conference Proceeding Series* (2010), pp. 195–198. DOI: 10.1145/1900441.1900476 (cited on page 6).

[26] Finn Kensing and Jeanette Blomberg. "Participatory Design: Issues and Concerns". In: *Computer Supported Cooperative Work* 7 (1988), pp. 167–185. DOI: 10.1023/A (cited on page 6).

[27] R. Stallman. "The GNU Manifesto". In: *Dr. Dobb's Journal of Software Tools* 10.3 (1985), pp. 53–70 (cited on page 7).

[28] Eric Raymond. *The cathedral and the bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*. 1999, p. 279 (cited on pages 7, 8, 225, 226).

[29] Eric von Hippel and Georg von Krogh. "Open Source Software and the "Private-Collective" Innovation Model: Issues for Organization Science". In: *Organization Science* 14.2 (2003), pp. 209–223. DOI: 10.1287/orsc.14.2.209.14992 (cited on page 9).

[30] Eric von Hippel. "Cooperation between rivals: Informal know-how trading". In: *Research Policy* 16.6 (1987), pp. 291–302. DOI: 10.1016/0048-7333(87)90015-1 (cited on page 9).

[31] Red Hat Inc. *The open source way*. 2009. URL: https://opensource.com/open-source-way (visited on 07/14/2020) (cited on page 9).

[32] Lawrence Lessig. *Free Culture*. New York: The Penguin Press, 2004 (cited on page 9).

[33] Jeff Howe. "The Rise of Crowdsourcing". In: *Wired Magazine* 14 (2006), pp. 1–5 (cited on pages 9, 13).

[34] Dennis Kaspori. "A Communism of ideas Towards an open-source architectural practice". In: *Archis* 3 (2003) (cited on pages 10, 12).

[35] Carlo Ratti et al. "Open Source Architecture". In: *Domus* (2011) (cited on page 10).

[36] Carlo Ratti and Matthew Claudel. *Open Source Architecture*. 1st ed. Thames & Hudson, 2015, p. 144 (cited on pages 10, 224, 226).

[37] Mark Garcia. "Architectural Patents and Open-Source Architectures: The Globalisation of Spatial Design Innovations (or Learning from 'E99')". In: *Architectural Design* 86.5 (2016), pp. 92–99. DOI: 10.1002/ad.2094 (cited on page 11).

[38] Aaron Sprecher and Chandler Ahrens. "Adaptive Knowledge in Architecture: A Few Notes on the Nature of Transdisciplinarity". In: *Architectural Design* 86.5 (2016), pp. 26–35. DOI: 10.1002/ad.2085 (cited on page 11).

[39] Wendy W Fok. “Opening Up the Future of Open Source: From Open Innovation to the Internet of Things for the Built Environment”. In: *Architectural Design* 86.5 (2016), pp. 116–125. DOI: 10.1002/ad.2097 (cited on page 11).

[40] Jonathan Dortheimer and Talia Margalit. “Open-source architecture and questions of intellectual property, tacit knowledge, and liability”. In: *Journal of Architecture* 25.3 (2020), pp. 276–294. DOI: 10.1080/13602365.2020.1758950 (cited on pages 11, 12, 33, 220, 224).

[41] Richard Stallman. *The Free Universal Encyclopedia and Learning Resource*. 2000. URL: https://www.gnu.org/encyclopedia/anencyc.txt (cited on page 12).

[42] Richard Stallman. *The Free Encyclopedia Project*. 2014. URL: https://www.gnu.org/encyclopedia/encyclopedia.html (cited on page 12).

[43] Mahmood Hosseini et al. “Crowdsourcing: A taxonomy and systematic mapping study”. In: *Computer Science Review* 17 (2015), pp. 43–69. DOI: 10.1016/j.cosrev.2015.05.001 (cited on page 13).

[44] Bruce Perens. “The Open Source Definition”. 2008 (cited on page 13).

[45] Tova Milo. “Crowd-Based Data Sourcing”. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 7108 LNCS (2011). Ed. by Shinji Kikuchi et al., pp. 64–67. DOI: 10.1007/978-3-642-25731-5_6 (cited on page 13).

[46] Maria Angelico and Imdat As. “Crowdsourcing Architecture : a Disruptive Model in Architectural Practice”. In: *ACADIA*. 2007. San Francisco: ACADIA, 2012, pp. 439–443 (cited on pages 13, 214).

[47] Pedro Miguel Hernandez Salvador Guilherme. “Competitions serve a larger purpose in architectural knowledge”. In: *Proceedings of the four International Conference on Architectural Research by Design*. Vol. 11. 2014, pp. 425–451 (cited on page 14).

[48] Magnus Rönn. “Judgment in the Architectural Competition – rules, policies and dilemmas”. In: *Nordic Journal of Architectural Research* 21.2-3 (2009), pp. 52–66 (cited on page 14).

[49] Peggy Deamer. “The Guggenheim Helsinki Competition: What Is the Value Proposition?” In: *Avery Review* 8 (2015), pp. 1–5 (cited on pages 14, 50, 222).

[50] Elizabeth Keslacy. “Arcbazar and the Ethics of Crowdsourcing Architecture”. In: *Thresholds* 46 (2018), pp. 300–317. DOI: 10.1162/thld_a_00043 (cited on pages 14, 50, 222).

[51] Nigel Cross. *The Automated Architect*. Pion Limited, 1977, p. 178 (cited on page 22).

[52] J K Page. “A review of the papers presented at the conference”. In: *Design Methods*. Ed. by J C Jones and D Thornley. Pergamon Press, Oxford, 1963, pp. 205–215 (cited on page 22).

[53] Christopher J Jones. “A method of systemic design”. In: *Design Methods*. Ed. by Christopher j Jones and D G Thornley. Pergamon Press, Oxford, 1963, pp. 53–73 (cited on pages 22, 214, 218).

[54] John Luckman. “An Approach to the Management of Design”. In: *OR* 18.4 (1967), p. 345. DOI: 10.2307/3007686 (cited on pages 22, 215).

[55] T A Markus. “The role of building performance measurement and appraisal in design method”. In: *The Architects' Journal* 20 December (1967), pp. 1567–1573 (cited on pages 22, 23).

[56] Stephen J Kline. “Innovation Is Not a Linear Process”. In: *Research Management* 28.4 (1985), pp. 36–45. DOI: 10.1080/00345334.1985.11756910 (cited on pages 23, 215).

[57] Robert P. Smith and Primanata Tjandra. “Experimental Observation of Iteration in Engineering Design”. In: *Research in Engineering Design - Theory, Applications, and Concurrent Engineering* 10.2 (1998), pp. 107–117. DOI: 10.1007/BF01616691 (cited on page 23).

[58] Dave Corne, Tim Smithers, and Peter Ross. “Solving design problems by computational exploration”. In: *IFIP WG5.2 on Formal Design Methods for CAD* 682 (1994), pp. 249–270 (cited on pages 23, 24, 214).

[59] Mary Lou Maher and Hsien Hui Tang. “Co-evolution as a computational and cognitive model of design”. In: *Research in Engineering Design* 14.1 (2003), pp. 47–64. DOI: 10.1007/s00163-002-0016-y (cited on page 24).

[60] Stefan Wiltschnig, Bo T. Christensen, and Linden J. Ball. “Collaborative problem-solution co-evolution in creative design”. In: *Design Studies* 34.5 (2013), pp. 515–542. DOI: 10.1016/j.destud.2013.01.002 (cited on page 24).

[61] Simon Austin, Andrew Baldwin, and Andrew Newton. “A Data Flow Model to Plan and Manage the Building Design Process”. In: *Journal of Engineering Design* 7.1 (1996), pp. 3–25. DOI: 10.1080/09544829608907924 (cited on page 26).

[62] Geoffrey Broadbent. *Design in Architecture: Architecture and the Human Science*. David Fulton Publishers Ltd., 1988, p. 504 (cited on page 26).

[63] RIBA. *RIBA Plan of Work 2013*. 2013. URL: https://www.ribaplanofwork.com/ (visited on 07/14/2020) (cited on pages 26, 202).

[64] Leon Batista Alberti and Giacomo Leoni. *The architecture of Leon Batista Alberti in Ten Books*. E. Owen, 1755, pp. 1–794 (cited on page 28).

[65] Neil Leach. “The Culture of the Copy”. In: *Architectural Design* 86.5 (2016), pp. 126–133. DOI: 10.1002/ad.2098 (cited on page 29).

[66] Richard Dawkins. *The Selfish Gene*. Oxford University Press, 1976, p. 224 (cited on page 29).

[67] Clare Olsen and SinÉad Mac Namara. *Collaborations in architecture and engineering*. Routledge, 2014, pp. 1–213 (cited on page 29).

[68] Chuck Eastman et al. *BIM Handbook: A Guide to Building Information Modeling for Owners, Managers, Designers, Engineers and Contractors*. 2nd ed. John Wiley & Sons, Inc., 2011, pp. 1–650 (cited on page 29).

[69] John Haymaker et al. "Filter mediated design: Generating coherence in collaborative design". In: *Design Studies* 21.2 (2000), pp. 205–220. DOI: 10.1016/S0142-694X(99)00042-3 (cited on page 29).

[70] Christopher Alexander et al. *A Pattern Language: Towns, Buildings, Construction*. Vol. 14. 1. Oxford University Press, 1977, p. 1216 (cited on page 30).

[71] Nikos A Salingaros et al. *P2P Urbanism*. Ed. by Nikos A Salingaros. Draft vers. Solingen: Umbau Verlag, 2011, pp. 1–116 (cited on page 30).

[72] N.John Habraken and Mark D. Gross. "Concept design games". In: *Design Studies* 9.3 (1988), pp. 150–158. DOI: 10.1016/0142-694X(88)90044-0 (cited on page 30).

[73] John Forester. *The deliberative practitioner: Encouraging participatory planning processes*. MIT Press, 1999, p. 305 (cited on page 30).

[74] Christopher Alexander et al. *The Oregon Experiment*. Oxford University Press, 1975, p. 202 (cited on page 30).

[75] Toni Robertson and Jesper Simonsen. "Challenges and Opportunities in Contemporary Participatory Design". In: *Design Issues* 28.3 (2012), pp. 3–9. DOI: 10.1162/DESI_a_00157 (cited on pages 30, 221).

[76] Michael Polanyi. *Personal Knowledge: Towards a Post-Critical Philosophy*. London: Routledge & Kegan Paul Ltd, 1958, p. 493 (cited on page 31).

[77] Michael Polanyi. *The Tacit Dimension*. University Of Chicago Press, 1966, p. 128 (cited on page 31).

[78] Ikujiro Nonaka. "A Dynamic Theory of Organizational Knowledge Creation". In: *Organization Science* 5.1 (1994), pp. 14–37. DOI: 10.1287/orsc.5.1.14 (cited on pages 31, 32).

[79] Tom Shaked, Karen Lee Bar-sinai, and Aaron Sprecher. "Autonomous in craft". In: *Anthropocene, Proceedings ofthe 25th International Conference ofthe Association for Computer-Aided Architectural Design Research in Asia (CAADRIA)*. Vol. 2. 2020, pp. 243–252 (cited on page 31).

[80] Gabriela Goldschmidt, Hagay Hochman, and Itay Dafni. “The design studio crit: Teacher-student communication”. In: *Artificial Intelligence for Engineering Design, Analysis and Manufacturing: AIEDAM* 24.3 (2010), pp. 285–302. DOI: 10.1017/S089006041000020X (cited on page 32).

[81] Aniket Kittur et al. “The Future of Crowd Work”. In: *Proc. CSCW ’13*. San Antonio, 2013, pp. 1–17. DOI: 10.1145/2441776.2441923 (cited on page 33).

[82] Daniela Retelny et al. “Expert Crowdsourcing with Flash Teams”. In: *UIST 2014 - Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*. Honolulu, HI, USA: ACM, 2014, pp. 75–86. DOI: 10.1145/2642918.2647409 (cited on page 33).

[83] Max Goldman, Greg Little, and Robert C Miller. “Real-time collaborative coding in a web IDE”. In: *24th annual ACM symposium on User interface software and technology (UIST ’11)*. New York, NY, USA: Association for Computing Machinery (ACM) Author’s, 2011, pp. 155–164. DOI: 10.1145/2047196.2047215 (cited on page 34).

[84] Thomas D LaToza et al. “Microtask programming: building software with a crowd”. In: *UIST ’14: Proceedings of the 27th annual ACM symposium on User interface software and technology*. 2014, pp. 43–54. DOI: 10.1145/2642918.2647349 (cited on pages 34, 39, 53).

[85] Mira Dontcheva, Elizabeth Gerber, and Sheena Lewis. “Crowdsourcing and creativity”. In: *Chi 2011*. Vancouver, BC, Canada: ACM, 2011, pp. 1–4 (cited on page 34).

[86] Aniket Kittur, Boris Smus, and Robert Kraut. “CrowdForge Crowdsourcing Complex Work”. In: *Human factors in computing systems*. Santa Barbara, CA, USA: ACM, 2011, pp. 43–52. DOI: 10.1145/1979742.1979902 (cited on pages 35, 38, 53, 214).

[87] Lingyun Sun et al. “Collaborative sketching in crowdsourcing design: a new method for idea generation”. In: *International Journal of Technology and Design Education* 25.3 (2015), pp. 409–427. DOI: 10.1007/s10798-014-9283-y (cited on pages 35, 53, 54, 74, 211).

[88] Lixiu Yu and Jeffrey V Nickerson. “Cooks or Cobblers? Crowd Creativity through Combination”. In: *Proceedings of the 29th Annual ACM SIGCHI Conference on Human Factors in Computing Systems*. Vancouver, BC, Canada: ACM, 2011, pp. 1393–1402. DOI: 10.1145/1978942.1979147 (cited on pages 36, 37, 53, 54).

[89] Hao Wu, Jonathan Corney, and Michael Grant. “Crowdsourcing Measures of Design Quality”. In: *Proceedings of the ASME 2014 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference; IDETC/CIE 2014*. August 2014. Buffalo, New York, USA CROWDSOURCING: ASME, 2014, pp. 1–10. DOI: 10.1115/DETC201434967 (cited on pages 37, 53).

[90] Aniket Kittur, Ed H. EH Chi, and Bongwon Suh. “Crowdsourcing user studies with Mechanical Turk”. In: *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08* November 2016 (2008), pp. 453–456. DOI: 10.1145/1357054.1357127 (cited on page 38).

[91] Nihar B. Shah and Dengyong Zhou. “Double or Nothing: Multiplicative Incentive Mechanisms for Crowdsourcing”. In: *Journal of Machine Learning Research* 17 (2014), pp. 1–9 (cited on page 38).

[92] Dana Chandler and Adam Kapelner. “Breaking monotony with meaning: Motivation in crowdsourcing markets”. In: *Journal of Economic Behavior & Organization* 90 (2013), pp. 123–133. DOI: 10.1016/j.jebo.2013.03.003 (cited on page 38).

[93] Aaron Shaw, John J. Horton, and Daniel L. Chen. “Designing Incentives for Inexpert Human Raters”. In: *CSCW2011*. Hangzhou, China: ACM, 2011. DOI: 10.2139/ssrn.1724518 (cited on page 39).

[94] Joel O. Wooten and Karl T. Ulrich. “Idea Generation and the Role of Feedback: Evidence from Field Experiments with Innovation Tournaments”. In: *Production and Operations Management* 26.1 (2017), pp. 80–99. DOI: 10.1111/poms.12613 (cited on page 39).

[95] Michael D. Greenberg, Matthew W. Easterday, and Elizabeth M. Gerber. “Critiki: A scaffolded approach to gathering design feedback from paid crowdworkers”. In: *C and C 2015 - Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition*. 2015, pp. 235–244. DOI: 10.1145/2757226.2757249 (cited on pages 39, 40).

[96] Kurt Luther et al. “Structuring, Aggregating, and Evaluating Crowdsourced Design Critique”. In: *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. New York, NY, USA: ACM, 2015, pp. 473–485. DOI: 10.1145/2675133.2675283 (cited on pages 39, 54, 59).

[97] Hillel Schocken. “Intimate Anonymity”. In: *Intimate Anonymity*. 2000, pp. 20–59 (cited on page 41).

[98] Hao Wu, Jonathan Corney, and Michael Grant. “An evaluation methodology for crowdsourced design”. In: *Advanced Engineering Informatics* 29.4 (2015), pp. 775–786. DOI: 10.1016/j.aei.2015.09.005 (cited on pages 53, 209).

[99] T. J. Howard, S. J. Culley, and E. Dekoninck. "Describing the creative design process by the integration of engineering design and cognitive psychology literature". In: *Design Studies* 29.2 (2008), pp. 160–180. DOI: 10.1016/j.destud.2008.01.001 (cited on page 54).

[100] Heinz Schmitz and Ioanna Lykourentzou. "Online Sequencing of Non-Decomposable Macrotasks in Expert Crowdsourcing". In: *ACM Transactions on Social Computing* 1.1 (2018), pp. 1–33. DOI: 10.1145/3140459 (cited on page 54).

[101] Christopher Frayling. "Research in Art and Design". In: *Royal College of Art Research Papers* 1.1 (1993), pp. 1–5 (cited on page 60).

[102] Nigel Cross. "Designerly Ways of Knowing : Design Discipline". In: *Design Issues* 4.3 (1982), pp. 221–227 (cited on page 60).

[103] Nigel Cross. "Designerly Ways of Knowing: Design Discipline Versus Design Science". In: *Design Issues* 17.3 (2001), pp. 49–55. DOI: 10.1162/074793601750357196 (cited on page 60).

[104] John Zimmerman, Erik Stolterman, and Jodi Forlizzi. "An analysis and critique of Research through Design". In: *Proceedings of the 8th ACM Conference on Designing Interactive Systems - DIS '10*. New York, New York, USA: ACM Press, 2010, p. 310. DOI: 10.1145/1858171.1858228 (cited on page 60).

[105] John Zimmerman, Jodi Forlizzi, and Shelley Evenson. "Research Through Design as a Method for Interaction Design Research in HCI". In: *Chi 2007*. San Jose, California, USA: ACM, 2007 (cited on page 60).

[106] Darren Gergle and Desney S. Tan. "Experimental Research in HCI". In: *Ways of Knowing in HCI*. New York, NY: Springer New York, 2014, pp. 191–227. DOI: 10.1007/978-1-4939-0378-8_9 (cited on page 62).

[107] Gabriela Goldschmidt. "Serial sketching: visual problem solving in designing". In: *Cybernetics and Systems* 23.2 (1992), pp. 191–219. DOI: 10.1080/01969729208927457 (cited on page 193).

[108] Daren. C. Brabham. "Crowdsourcing the Public Participation Process for Planning Projects". In: *Planning Theory* 8.3 (2009), pp. 242–262. DOI: 10.1177/1473095209104824 (cited on page 221).

[109] Hamid R. Ekbia and Bonnie Nardi. *Heteromation, and Other Stories of Computing and Capitalism*. Cambridge, Massachusetts: The MIT Press, 2017, p. 266 (cited on page 222).

[110] Matthew Fuller and Usman Haque. *Urban Versioning System 1.0*. 1st ed. New York, New York, USA, 2008 (cited on page 224).

[111] Anand Kulkarni, Matthew Can, and Björn Hartmann. "Turkomatic : Automatic Recursive Task and Workflow Design for Mechanical Turk". In: *Aaai* (2011), pp. 2053–2058. DOI: 10.1145/1979742.1979865 (cited on page 227).