

**A Method for Evaluating the Architectural Quality of Storefronts Using
Statistical Methods**

Jun Yu

A thesis

Submitted in partial fulfillment of the
requirement of the degree of

Master of Science in Architecture: Design Computing

University of Washington

2022

Committee:

Alex T Anderson

Alex Colburn

Program Authorized to Offer Degree:

Architecture

©Copyright 2022

Jun Yu

University of Washington

Abstract

A Method for Evaluating the Architectural Quality of Storefronts Using Statistical Methods

Jun Yu

Chair of the Supervisory Committee:

Alex T. Anderson

Department of Architecture

Due to the complexity and long-life cycle of architectural projects, the evaluation of architectural performance is an irreplaceable part of design. Existing systematic building evaluation methods mainly focus on structural, environmental, and economic factors, while ignoring the establishment of evaluation systems for subjective factors such as aesthetics and context. This is partly due to the multidisciplinary complexity of human behavior studies and partly due to the difficulty of quantifying subjective data. It leads to the fact that designers and design review boards have to pay extra time and capital costs to deal with these unclear criteria. This thesis aims to establish an architectural evaluation system based on qualitative and quantitative research into different features of architectural facades, such as color, geometry, etc. Focusing on one building type, small-scale commercial storefronts, and using statistical methods, we build mathematical models to describe and predict people's aesthetic preferences for key design criteria. We plan to leverage quantitative methods to help establish a system for evaluating aesthetic choices that design review boards can use to make their decisions more consistent across projects and jurisdictions, and to fill the gap between them and designers.

Acknowledgments

Firstly, I would like to express my deepest appreciation to both my committee chair, Prof. Alex Anderson and committee member, Prof. Alex Colburn. With a profound understanding of architecture research, Alex Anderson has provided concise guidance on the orientation of the research and valuable suggestions of details throughout the process. Thanks to him, we could reach out to Prof. Jim Nicholls for expert advice, and to potential study subjects so we could collect survey responses quickly. During the research process, Alex Colburn contributed much more than technical help. As we are dealing with a complex system development problem, he always foresees the potential mistakes and provides useful feedback advice in advance. In addition, the work would not be possible without his guidance in statistical analysis. Both of them are incredible mentors and I am pleased that they were able to join my thesis committee.

I also want to say thank you to all my family members. They always offer most generous support to whatever decision I make. Their greetings and phone calls gave me strength to go through the whole process during research.

Finally, I would thank Prof. Mehlika, Prof. Thomas, Prof. Jim Nicholls, Prof. Brian Johnson, my friends and roommates – whoever provided valuable advice on or outside of the research. I am proud and feel very fortunate to have met them in my life. This will be a precious memory that I never forget.

Table of Contents

1. Introduction	1
2. Background – Finding in architecture evaluation methods	2
2.1 Multiple-criteria decision analysis	2
2.2 Design review boards (DRB)	4
2.2.1 Evaluation methods	4
2.2.2 Problems	5
2.3 Geometry analysis	6
2.4 Color harmony	9
2.5 Context	11
3. Methodology	13
3.1 Workflow introduction	13
3.2 Concept development	14
3.3 Qualitative research	14
3.3.1 Expert interview	14
3.3.2 Site photos collection and semi-structured interview	15
3.3.3 Iteration of refining features	17
3.4 Quantitative research	19
3.4.1 Mechanism of survey	19
3.4.2 Regularity	22
3.4.3 Color combination	23
3.4.4 Overall evaluation	23
3.4.5 Context	24
3.5 Statistical analysis	24
4. Results	26
4.1 Paired test to multiple data sources	26
4.2 Reliability test	28
4.3 Normal distribution test	29
4.4 Correlation analysis	29
4.5 Relationship between regularity score and likability	31
4.6 Relationship between color combination score and likability	34
4.7 Relationship between overall score, regularity, and color combination	37
5. Discussion	39
6. Conclusion	39
Reference	41
Appendix A: Questionnaire A	43
Appendix B: Questionnaire B	47
Appendix C: Paired Wilcoxon Test	53
Appendix D: Result of Shapiro Wilk test	56

1.Introduction

Evaluation of architectural design is often considered to be highly complex and difficult because it covers not only objective criteria such as structure, energy efficiency, land use, functionality, and spatial form, but also subjective criteria such as aesthetic choices, and context integration. Systematic methods, such as the Cost Benefit Analysis Method (CBAM) (Kazman et al. 2002), which are based on objective factors analysis have been developed for the evaluation of some of these areas. However, a systematic framework for subjective evaluation of architecture has not yet been developed. Although research on cognition and perception that is relevant to architectural evaluation has been carried out in the field of human behavior study, most of these studies focus on issues that are only generally related to architecture such as color harmony (Chamaret and Urban,2013), symmetry (Malewczyk et al. 2022), emotion (Janssen et al. 2017), etc. These features are studied individually and lack the integration needed to apply their findings effectively in architecture.

As the group most frequently tasked with architectural evaluation of aesthetic factors, a municipality's design review board is typically responsible for reviewing site design, architectural design, and landscaping for all public and most private development applications to make sure they meet the land use code. However, local rules about aesthetics and context are often vague and imprecise. Furthermore, designers tend to rely on personal experience in evaluating architectural color and style (Wang, 2013), which increases the risk of being rejected by the design review board. This can cause a significant loss of investment in time and capital costs.

The objective of this thesis is to regularize the criteria for analysis of architectural facades to a certain extent by extracting and evaluating some vital visual features. Its goal is to make the review/ self-examination process quantitative, measurable, and communicable.

This thesis involved qualitative research and statistical analysis. It included an interview with an expert on small-scale retail design, study of real-world storefronts, a pre-test with non-experts, and a survey with more than 120 test subjects. The survey focused on four aspects of architecture that are relevant to a design review board's evaluation of a building: geometric regularity, color combination, overall aesthetic preference, and relationship to context. Using statistical methods, we sought to determine the overall aesthetic score of chosen storefronts by analyzing features of regularity and color combination.

The primary contribution of this thesis is a preliminary study that explores the feasibility of utilizing visual features to quantify subjective criteria used by design review boards. It is a step towards filling the gap between designers and design review boards by integrating multiple subdivided features into an overall study and applying it to architecture. Understanding the open-ended nature of subjective evaluation, the goal is to provide a systematic framework to add clarity to the aesthetic judgements of design review boards.

2. Background

2.1 Multiple-criteria decision analysis

In many cases, the decisions in architectural design are so complicated that architects cannot rely solely on logic (Schmoldt et al, 2000). Despite empirical decision-making being acceptable in many simple design projects, complex architectural problems require more rigorous evaluation methods (Youdale, 1983).

With the development of computer science and statistical analysis, methods like Multi Criteria Decision Making (MCDM) and Decision Analysis (DA) tools have been applied in the architecture, engineering and design (AED) industry. They aim to support resolution of technical or economic issues, where efficiency tends to be more important than aesthetic judgment (Gololov and Yezioro, 2007). For example, a computerized support system for comparative evaluation of building technologies for building envelopes automates the stages of the MCDM procedure (see Fig. 2.1.1). It presents an organized approach for decision making techniques for defining priorities and defines common selection rules (Maguire, et.al, 1994). It helps to provide the guidance for choosing the most appropriate features to the methods.

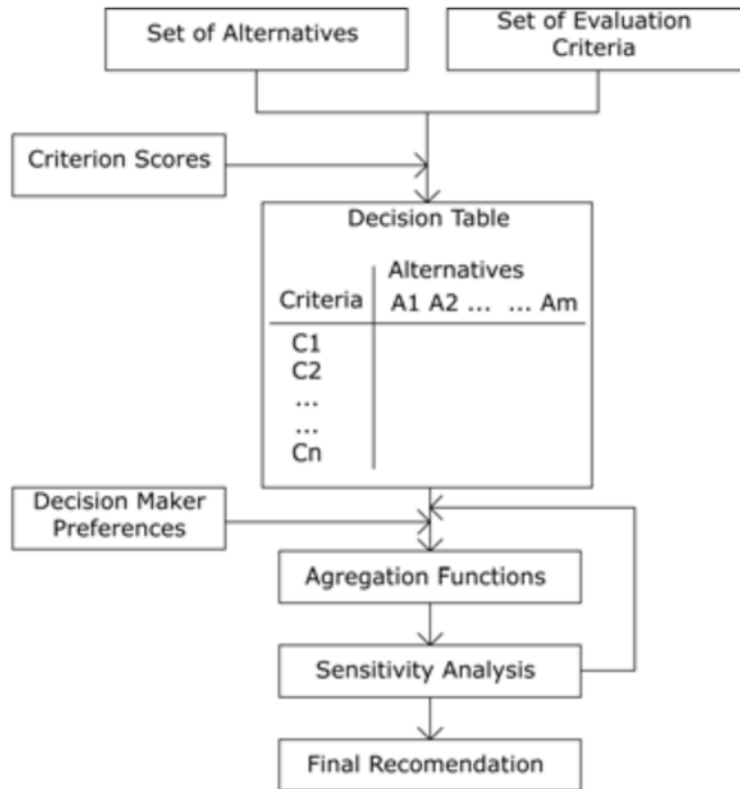


Figure 2.1.1 MCDM General Model, Gololov and Yezioro, A COMPUTER SYSTEM FOR MULTI-CRITERIA COMPARATIVE EVALUATION OF BUILDING ENVELOPES, Simulation, 2007

However, such quantitative and measurement methods are mainly focused on the cost management (Chigara et al. 2013) and energy analysis (Bakar et al. 2014). The two most widely used architecture evaluation methods, which were developed years ago are: the Architectural Tradeoff Analysis Method (ATAM) (Clements et al. 2002), which helps select a suitable structure for system development by discovering sensitive points; and the Cost Benefit Analysis Method (CBAM) (Kazman et al. 2002), which helps quantify and compare the total cost of undertaking the project.

Little research nowadays focuses on the human perceptual evaluation of architecture and quantifying people’s aesthetics preference towards it in a systematic way, partly because of the difficulty of accessing massive perceptual data which cannot easily be generated by computational simulation. In addition, due to the complex nature of aesthetics, figuring out general criteria for subjective evaluation is also a challenge.

This thesis takes the MCDM workflow as a reference and applies it to the evaluation of architectural façades based on a limited range of aesthetic criteria. From this study, we plan to build a systematic model for aesthetic evaluation of architecture that accounts

for the interaction of diverse building features (geometry, color, relationship to context, etc.)

2.2 Design review boards (DRB)

2.2.1 Evaluation methods

The design review board (DRB) in a municipality is a typically a “volunteer body that is primarily responsible for reviewing site design, architectural design and landscaping for all public and private development applications except single family residences” (Langley Washington, [www.langleywa.org/government/citizen_boards/design_review_board_\(drb\).php](http://www.langleywa.org/government/citizen_boards/design_review_board_(drb).php)). Members of the DRB consider a broad set of design considerations that the architect must use to design the exterior of a building. Their guidance and recommendations help to promote designs that adhere to local codes and that fit into and relate to surrounding neighborhoods. They also offer flexibility with the code standards to achieve better designs.

In Seattle, where the research for this thesis was conducted, the Seattle Department of Construction and Inspections (SDCI), states that the DRB reviews projects at public meetings at least twice – at an Early Design Guidance phase and a Master Use Plan phase. (SDCI website, <https://www.seattle.gov/sdci/>)

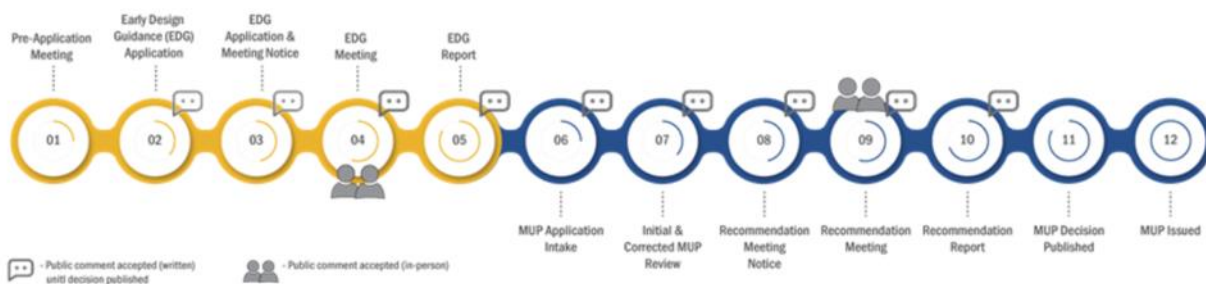
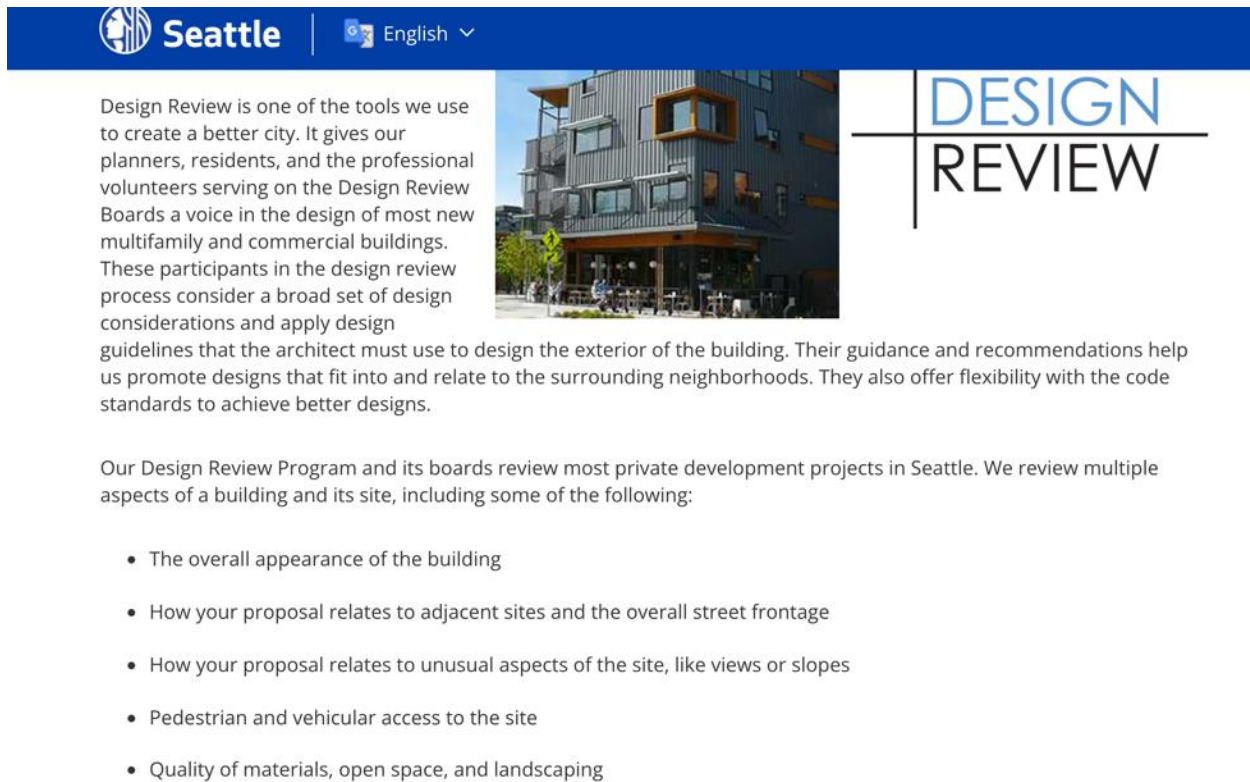


Figure 2.2.1 Workflow of DRB (Credit: Seattle for Everyone, Share the Cities Action Fund, and AIA Seattle), Reprinted from (<https://www.theurbanist.org/2021/09/22/fixing-seattles-design-review-process/>) used without permission

At the Early Design Guidance phase, the DRB reviews the project site, proposed development, and early design concepts for the design. For the Master Use Plan evaluation, the DRB conducts reviews to make sure the design meets the land use code (SDCI website, <https://www.seattle.gov/sdci/>); this includes evaluation of various building components, such as facades, height, cornice line, etc. In general, this is an iterative process, which means applicants must repeat the process if they cannot meet the recommendations from DRB. According to online data collected from 2010 to 2020

(Trumm, 2021), it takes more than 18 months to get through Master Use Permit and about 26 months to get a final permit (see Fig. 2.2.1).

2.2.2 Problems



Design Review is one of the tools we use to create a better city. It gives our planners, residents, and the professional volunteers serving on the Design Review Boards a voice in the design of most new multifamily and commercial buildings. These participants in the design review process consider a broad set of design considerations and apply design guidelines that the architect must use to design the exterior of the building. Their guidance and recommendations help us promote designs that fit into and relate to the surrounding neighborhoods. They also offer flexibility with the code standards to achieve better designs.

Our Design Review Program and its boards review most private development projects in Seattle. We review multiple aspects of a building and its site, including some of the following:

- The overall appearance of the building
- How your proposal relates to adjacent sites and the overall street frontage
- How your proposal relates to unusual aspects of the site, like views or slopes
- Pedestrian and vehicular access to the site
- Quality of materials, open space, and landscaping

Figure 2.2.2.1 Job description of DRB, Reprinted from (<https://www.seattle.gov/sdci/about-us>) used without permission

According to the Seattle DRB website (see Fig. 2.2.2), several of the evaluation criteria the board uses are subjective and vague to some extent, as they evaluate, for example, the “overall appearance of the building” and “how the proposal relates to adjacent sites.” (SCDI website, <https://www.seattle.gov/sdci/about-us/who-we-are/design-review>) Compared to codes that apply to specific architectural components or to district guidance, the evaluation of aesthetics and relationship to context are subjective and open ended. (See Fig. 2.2.3). There are no existing tools or methods that can provide quantitative references to DRBs to help them systematize their judgements on these subjective factors, leading to potential misunderstandings and disagreements between DRB and designers, and increased project approval time.

CS3 Architectural Context and Character: Contribute to the architectural character of the neighborhood.

CS3-A Emphasizing Positive Neighborhood Attributes

CS3-A-1. Fitting Old and New Together: Create compatibility between new projects, and existing architectural context, including historic and modern designs, through building articulation, scale and proportion, roof forms, detailing, fenestration, and/or the use of complementary materials.

CS3-A-2. Contemporary Design: Explore how contemporary designs can contribute to the development of attractive new forms and architectural styles; as expressed through use of new materials or other means.

CS3-A-3. Established Neighborhoods: In existing neighborhoods with a well-defined architectural character, site and design new structures to complement or be compatible with the architectural style and siting patterns of neighborhood buildings.

CS3-A-4. Evolving Neighborhoods: In neighborhoods where architectural character is evolving or otherwise in transition, explore ways for new development to establish a positive and desirable context for others to build upon in the future.

CS3-B Local History and Culture

CS3-B-1. Placemaking: Explore the history of the site and neighborhood as a potential placemaking opportunity. Look for historical and cultural significance, using neighborhood groups and archives as resources.

CS3-B-2. Historical/Cultural References: Reuse existing structures on the site where feasible as a means of incorporating historical or cultural elements into the new project.

Figure 2.2.2.2 Seattle Design Codes – Architectural Context and Character part, Reprinted from(<https://www.seattle.gov/documents/departments/opcd/vault/citywidedesignguidelinesupdate/seattledesignguidelines.pdf>) used without permission

Recent press articles suggest that there are significant problems with the Seattle DRB process (Natalie, 2021), including long delays and an unpredictable evaluation process. Design Review delays are growing, adding direct costs to building construction. This includes well-documented instances when high quality projects are held up for subjective reasons despite having broad community support. This process can also “push” clients and architects to develop “monochromatic projects” that are way too similar because of concerns about unknown or subjective criteria, which does harm to architectural diversity and creativity in the long run. (Trumm, 2021)

2.3 Geometry analysis

Geometrical transformation and stylization in the design process affect volumetric development of buildings and patterned organization of facades. Automated building technologies have the potential to systematize this, but they have also contributed to a growing interest in geometrical irregularity in building form. (Janssen et al. 2017)

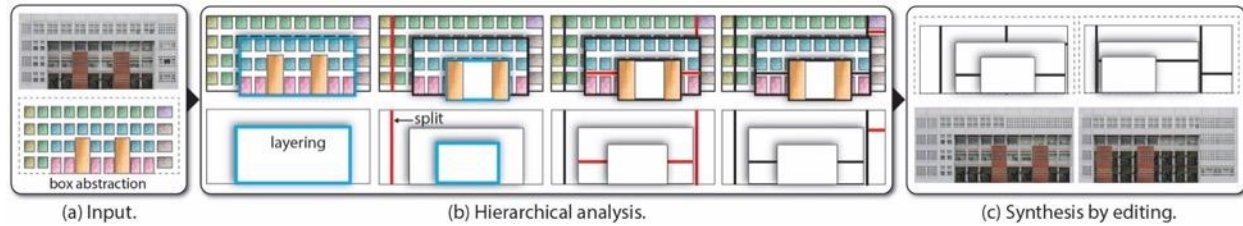


Figure 2.3.1 Overview of facade analysis and synthesis, Reprinted from (Zhang et al. Layered Analysis of Irregular Facades via Symmetry Maximization, [ACM Transactions on Graphics Volume 32 Issue 4](https://doi.org/10.1145/2461912.2461923), July 2013 Article No.: 121pp 1–13, <https://doi.org/10.1145/2461912.2461923>) used without permission

Existing façade analysis methods enable researchers to calculate the degree of irregularity on facades (Musialski et al. 2012). Some papers (Zhang et al. 2018) have developed algorithms for analyzing irregular 2D facades. Inspired by Gestalt Law of Prägnanz, which emphasizes the prevalence of symmetry and regularity in perceptual grouping (Wertheimer 1923), they recursively decompose facade elements based on measures of symmetry or repetition. Their research provides theoretical methods to quantify the degree of irregularity in facades, enabling us to determine the extent of regularity as a known design factor.

Recent research (Malewczyk et al. 2022) aims to determine the aesthetic preferences of people regarding the degree of regularity in the composition of building facades, and to identify these preferences depending on the type of geometric composition. This study included a questionnaire conducted with 109 people that asked them to react to the design of simplified building facades with various fenestration patterns. Statistical analyses of the obtained results were used to formulate answers to research questions, which included: “(1) Is there a relationship between the degree of regularity in facade compositions and aesthetic preferences? (2) Are regular compositions considered more aesthetic than those with a greater degree of irregularity? (3) Is there a relationship between the type of composition and aesthetic preferences? (4) What kind of composition is preferred the most or the least?” (Malewczyk et al. 2022, page 4) In the end the research found that people’s aesthetics preference is proportional to the degree of a façade’s regularity in general. Facades with a regular composition were found to be more aesthetically pleasing than those with an irregular composition.

Inspired by the work of Malewczk, we used a similar workflow to conduct the research for this thesis, including proposing hypotheses, survey design and conduct, and statistical analysis to draw the conclusions. The difference is that we present respondents with realistic façade photos for evaluation, and we use regularity and other features to access an overall aesthetics evaluation and study the correlation between diverse variables systematically.

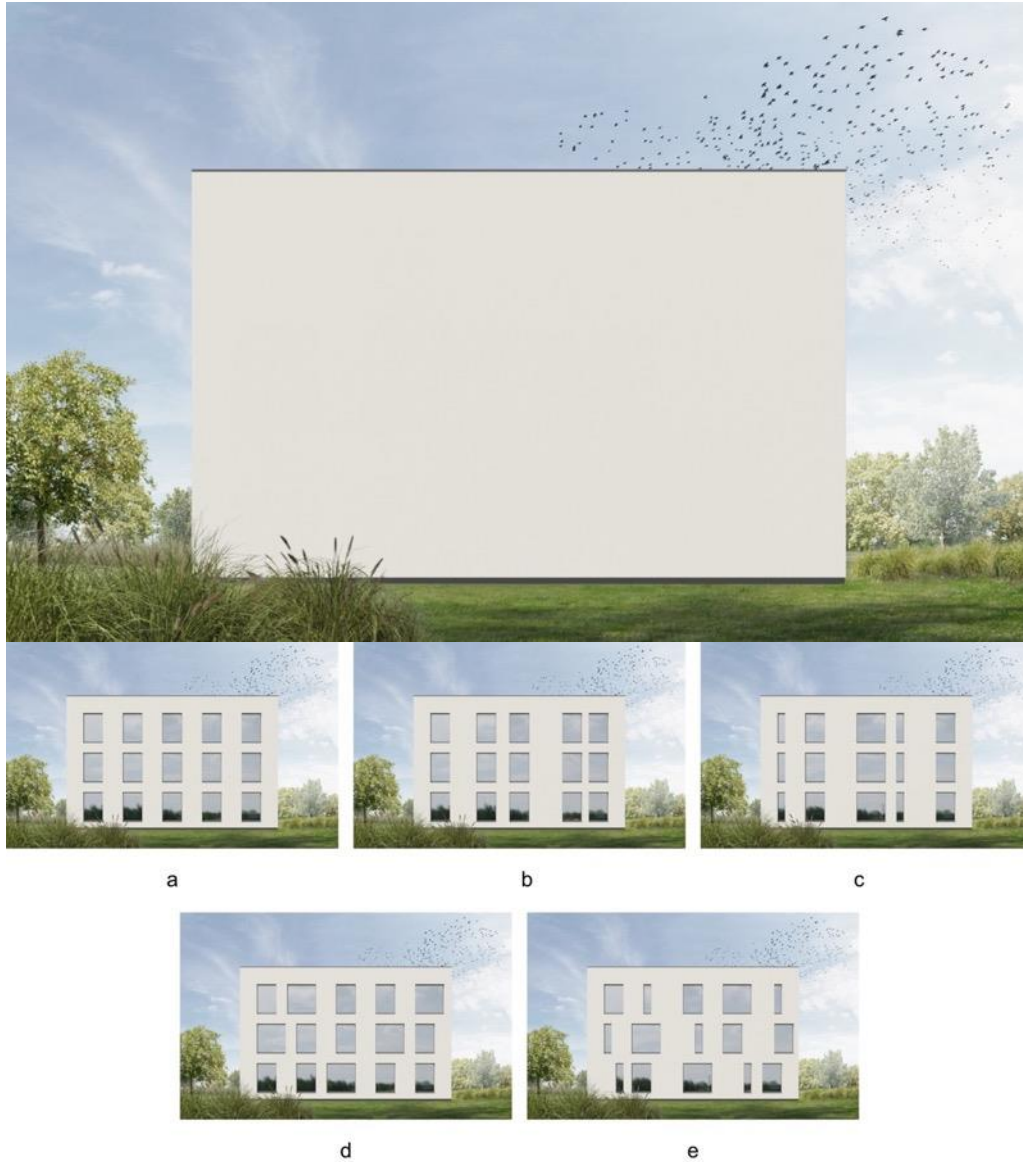


Figure 2.3.2 The incentives presented in the study, Reprinted from (Malewczyk, M.; Taraszkiewicz, A.; Czyż, P. Preferences of the Facade Composition in the Context of Its Regularity and Irregularity. *Buildings* 2022,12,169. <https://doi.org/10.3390/buildings12020169>) Used without permission

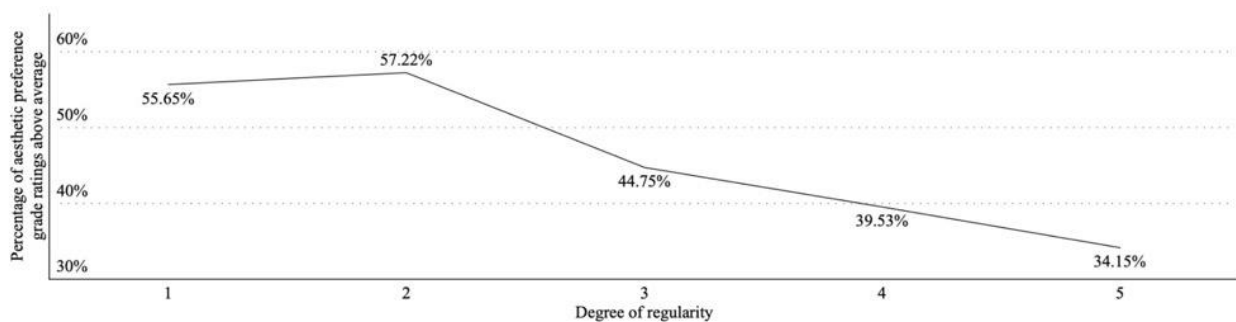


Figure 2.3.3 Percentage of aesthetic preference grade above average, Reprinted

from (Malewczyk, M.; Taraszkiewicz, A.; Czyż, P. *Preferences of the Facade Composition in the Context of Its Regularity and Irregularity. Buildings* 2022,12,169. <https://doi.org/10.3390/buildings12020169>
) Used without permission

2.4 Color harmony

Harmony of simple color patterns has been widely studied for color design. Rules defined by psychological experiments have been applied to derive aesthetic scores for images, or to automate re-coloration of pictures. (Chamaret and Urban, 2022) Researchers point out that “harmony” is a subjective concept whose difficulties “lie on its definition and its measurement.” Fedorovskaya et al. conducted a series of experiments to identify image features that create harmony (dis)comfort. They found out that edge contrast, average lightness, and range of lightness may influence global harmony appreciation. Other harmony investigations in image processing tend to focus on color combinations and complementarity. Recently, Moon and Spencer have performed psychological experiments to evaluate the relationship between color patterns and feeling feedback on different subjects. They have developed a two-color harmony model (1994). “Matsuda extended this model with contrast measurement where harmonious doublets and triplets of color are defined. Through his experiments, he defined HSV (Hue, Saturation, Value)-based templates that predict sets of harmonious hues. “(Chamaret and Urban, 2022) The results of these experiments are primarily applicable in graphic and fashion design. However, these principles also provide potential for evaluating color harmony in architecture.

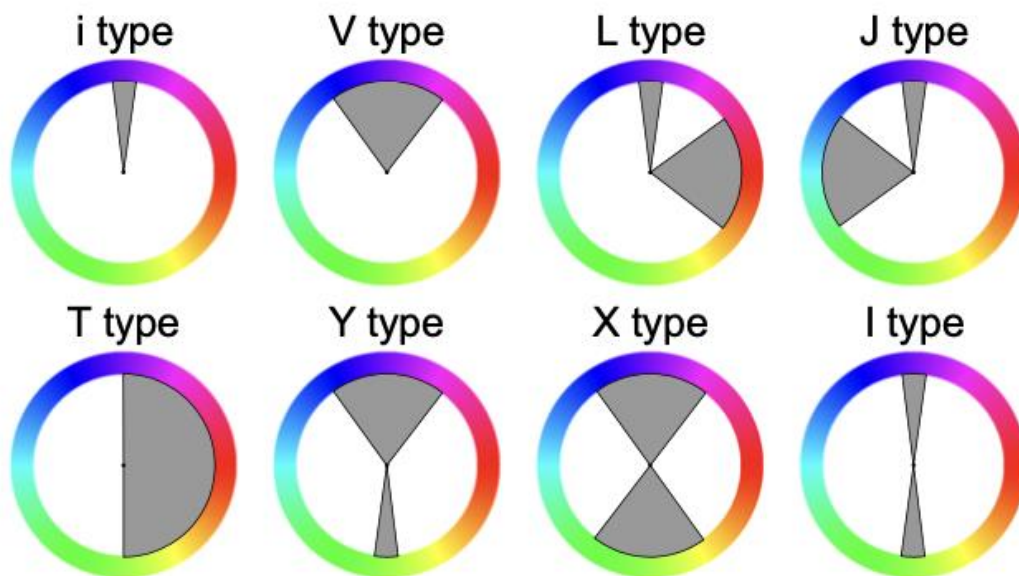


Figure 2.4.1 Harmonious hue templates defined by Matsuda experiments, Reprinted
1. from (Y. Matsuda. *Color design. Asakura Shoten, 1995*) Used without permission

In 2013, Chamaret and Urban proposed a new approach for assessing the quality of a picture. As a tool for content creators who want to maximize the aesthetic impact of their work, the proposed metric provides a perceptual harmony quality map as well as a score for disharmonies.

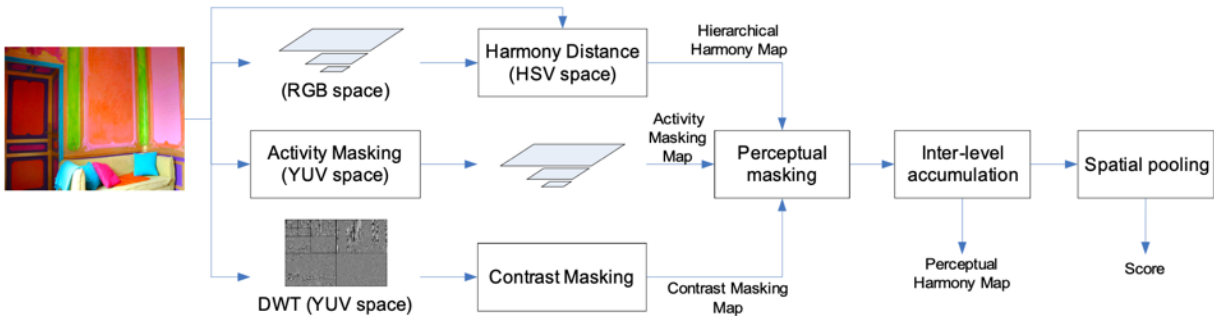



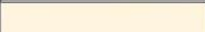
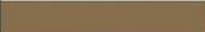
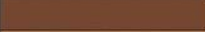



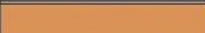

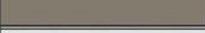
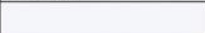
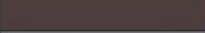



Figure 2.4.2 Overview of the complete system. Reprinted from (Chamaret and Urban, No-reference Harmony-guided Quality Assessment, CVPR2013) Used without permission

Other research also introduces color harmony into building exterior evaluation (Mehdipour ,2020). Mehdipour selected forty research publications published between 1960 and 2020 that discuss relationships between color and emotion. After examining the selected publications, they collected 97 color-emotion scales for color preference evaluation. They used the HSL system for the study because it provides numerical values of physical properties of a color, i.e., hue (H), saturation (S), and Luminance (L), which can be specified in this system. According to the obtained results, the researchers determined that the main components that contribute to the harmony-contrast of color combinations are the difference in luminance, the difference in saturation, and the temperature difference of the hues of the color combination in building exterior. These results confirm the harmony/contrast classification of color combination in the study carried out by Sarica and Cubukcu (2018).

Colour	H	S	L	Area
B001				
	25	55	89	48%
	25	48	77	33%
	32	35	49	19%
B003				
	41	13	100	40%
	36	44	53	25%
	19	59	46	18%
	27	45	76	17%
B025				
	17	61	52	71%
	34	21	99	29%
B029				
	28	60	85	61%
	20	65	71	36%
	42	16	51	3%
B054				
	228	2	98	56%
	3	22	30	39%
	210	8	51	5%

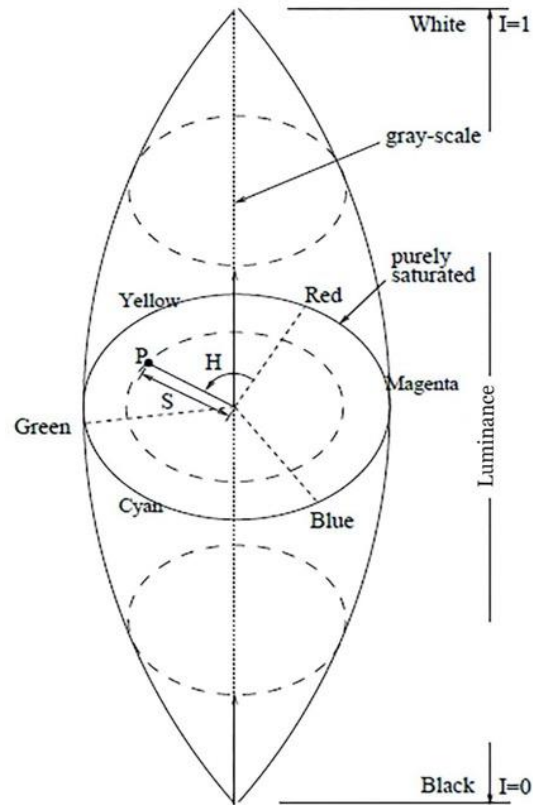


Figure 2.4.3 Left: The HSL codes of the colors of selected images that were related to the difference in saturation, Reprinted from (Mehdipour et al. 2021) Used without permission Right: HSL color system. Reprinted from (Citation: Sarica, C., & Cubukcu, E. 2018) Used without permission

2.5 Context

In an architectural sense, context gives meaning to parts of a building by reference to its surroundings. The context of a building includes physical/natural factors, socio-cultural factors (for example, the site's previous use) and so on. Assessment of context gives an idea of what is existing before a designer's plan to introduce a new component to it. According to Thomas and Garnham (2007), one of the most crucial questions for architecture and environmental design is "(page 7)" This complex question requires an equally rich response. The answer is not merely a quantitative one, but a qualitative one which gives full consideration to diverse emotions. "(Designing Buildings, <https://www.designingbuildings.co.uk/wiki/Context>)The context of a building may include:

- The topography of the area.
- The site's history and previous uses.
- Local culture.
- Architectural style.

- Local materials and construction techniques.
- Weather and microclimate.
- Political conditions.
- National and local policy.
- The state of the economy.



Figure 2.5 Example of context, Kyoto, Japan, 2017

Since the evaluation of context is multi-dimensional, it has been difficult to form a quantitative standard for this. In addition, calculating the weight of each feature also adds to the complexity of systematic framework. As a result, this thesis focuses on evaluating the general quality of building's contextual response by recording to the subjective evaluation of a large number of test subjects. We then use statistical analysis to help determine the degree to which the façade fits in its context.

3.Methodology

3.1 Workflow introduction

The research process involved four parts: concept development (background research, proposal statement), qualitative research (interviews and pre-test), quantitative research (experiment design and survey), and statistical analysis. Since the thesis is dealing with a systemic problem, some of these processes are iterated to address recurring sub questions.

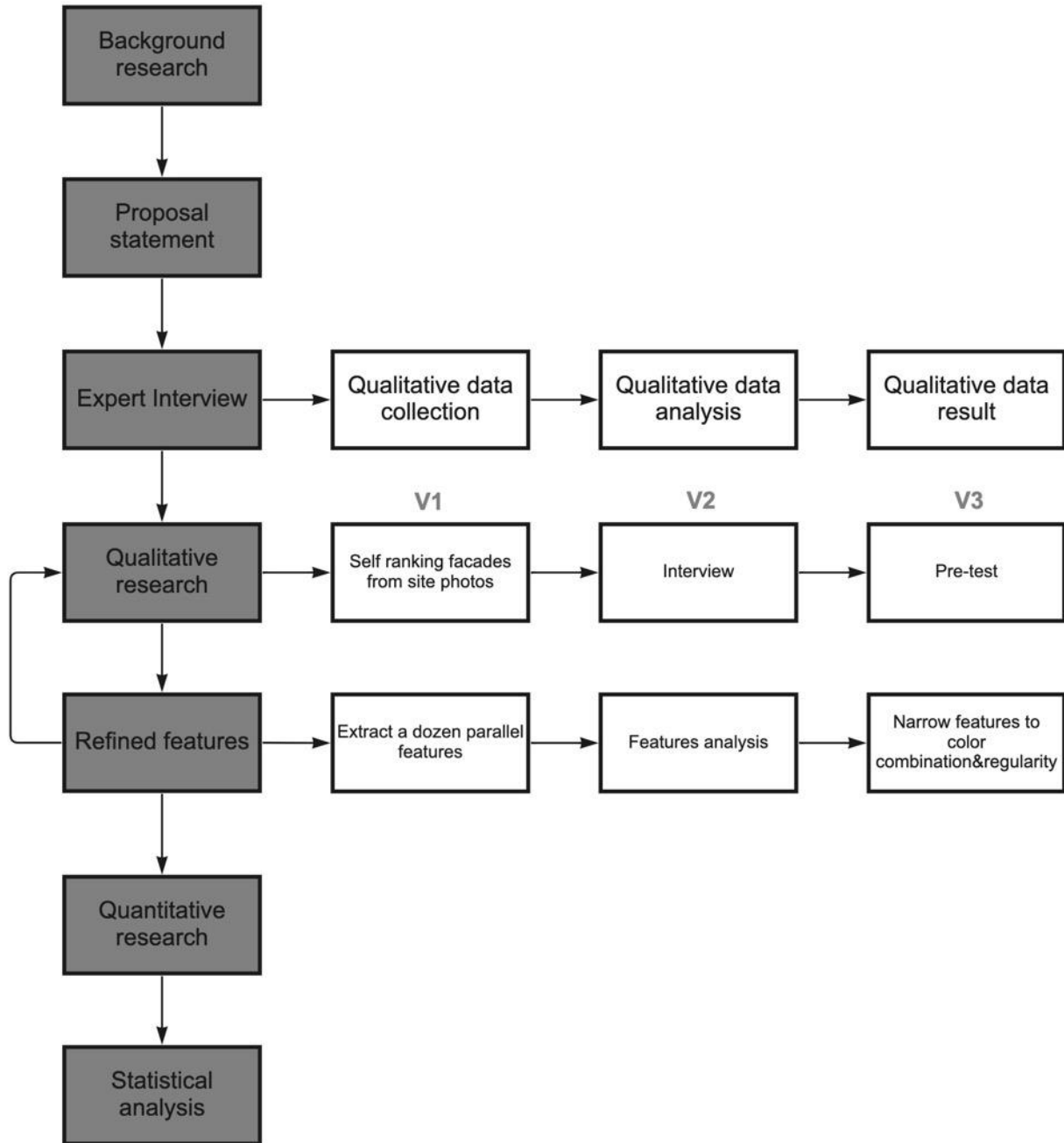


Figure 3.1.1 The research process

3.2 Concept development

The initial proposal of research is to study the styles of buildings in diverse Seattle districts and to simulate the styles based on deep learning methods. As we developed the study, examining the dominant features of facades in each district, we determined the more realistic problem for study is a lack of unified aesthetic principles in architectural evaluation. Then we adjusted the goal of research to help fix the gap between DRB and architects. The key component of the research design for the thesis is

exploring the workflow for developing an evaluation system, rather than merely focusing on data-driven analysis.

3.3 Qualitative research

3.3.1 Expert interview

Early in the process of research we consulted Jim Nicholls from the Architecture Department at University of Washington, where he has run the Storefront Studio for almost 20 years. In that course, students are required to design a community street front based on community outreach, preservation, and context.

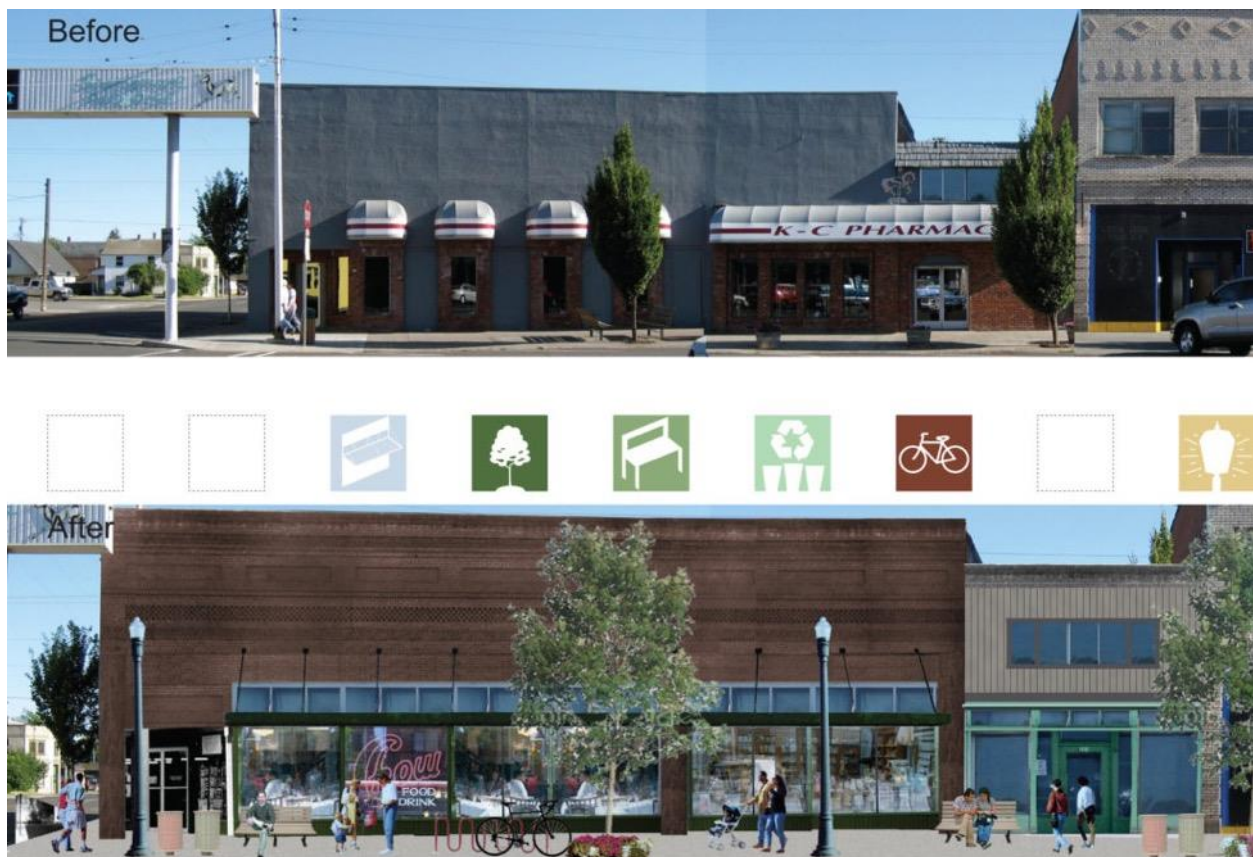


Figure 3.3.1 Storefront studio, Goldendale, WA, 2009, Reprinted from (<https://sfs.be.uw.edu/locations/about-goldendale/>) Used without permission

Prof. Nicholls shared his experience with storefront design and design review. Two major points emerged from the interview:

1. The focus of establishing an evaluation system criteria should be to find suitable features that can be judged using automated procedures. A feasible way to begin doing this is to field visit different storefronts and manually score them. At the same time, mark their features, such as the area occupied by the windows, the aspect ratio of the

facade, and so on. Summarize as many potential features as possible by comparing elevations and labels with different scores, although duplicates may be generated in the process.

2. Contextual design is a more complex task than individual façade design because different neighborhood may have diverse styles: historic district, commercial street, etc. Even blocks with similar typology may have subtle differences from each other. Therefore, it is important to first select several blocks with different styles to summarize their features and observe whether these features can replace each other. Define the features as precisely as possible to express the contextual fusion of the building. Nicholls offered in-depth insights into the roles of context and design process for storefronts. He also suggested some ways to develop the thesis research that were especially helpful in the photo survey and in developing the questionnaire described in section 5.3 below.

3.3.2 Site photos collection and Semi-structured interview

Based on the suggestions in 3.2.1, we collected about 220 facade photos from the following locations in Seattle.

- Ne 75th St (10)
- University District (124)
- Ravenna (15)
- University Village (68)



Figure 3.3.2 Ranked site photo, University Village, 2022

As is indicated in section 2.2.1, the Seattle design review board inspects multiple aspects of a building, including “the overall appearance of the building” (single building evaluation), “how the proposal relates to adjacent sites” (context evaluation).

Accordingly, for each façade we applied the following ranking criteria to:

- Façade quality
- Correspondence with surrounding buildings

For this evaluation, we used a scale from 1 to 10, with 1 indicating a very unattractive façade, or a façade that corresponds very poorly with surrounding buildings. By ranking and labeling building façades in this way we can begin to compare them and to understand potential correlations between different factors. For example, if two façades are otherwise similar except that the window to wall ratio is different, and their scores in this test are significantly different, then it can be reasonably concluded that the window to wall ratio is a feature that affects people's aesthetic evaluation of a façade.

For the context evaluation, this test revealed that it is useful to find the commonalities between the target facade and surrounding buildings and observe what features are most involved in the judgement of their correspondence. For example, is cornice height more or less important than surface material in determining whether a façade fits in its context?

After this initial façade ranking process, we conducted semi-structured interviews to test other people’s opinions about these façades. In the first round, two interviewees were asked to evaluate ten facades on the basis of the following criteria:

- The extent you like this façade (1-10, least to most)
- The context score of the façade (1-10, worst fit to best fit)
- Why do you like/dislike it? (spoken response)
- What feature affects this façade’s evaluation most? (spoken response)
- What feature do you think is most important for façade evaluation? And why? (spoken response)

This process helped us to determine which design features to have subjects evaluate in the more formal questionnaire.

3.3.3 Iteration of refining features

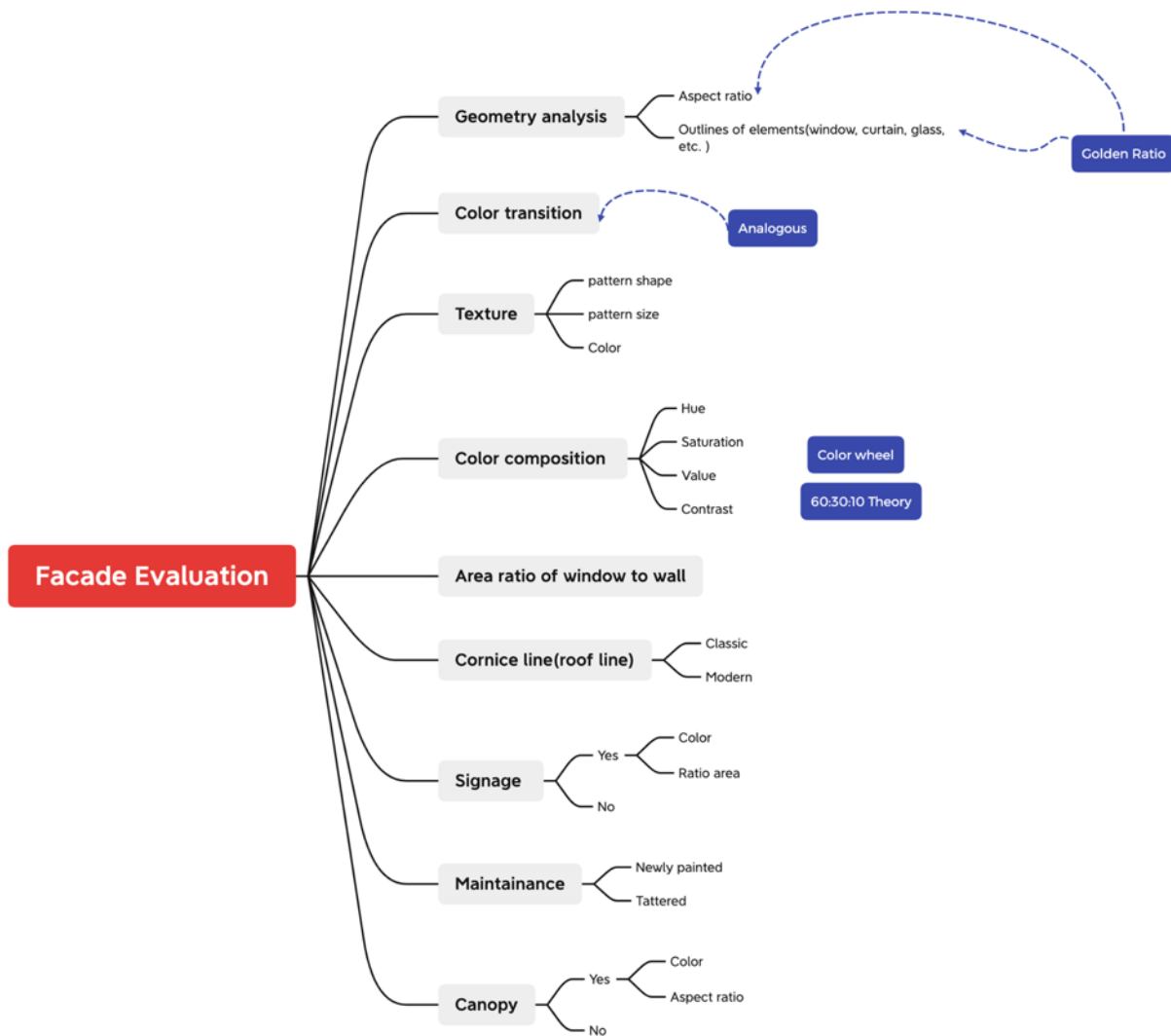


Figure 3.2.3 Features selected after first round iteration

Twelve features for possible evaluation arose in the photo analysis and the semi-structured interview; this includes several subdivisions under each category. Some of these features are somewhat redundant, for example, color transition & color composition and geometry analysis & texture are related to each other. The next goal, therefore, was to select informative features that are independent from each other. In order to realize this, we introduce the concept of information theory. This includes the concept of “entropy,” which can act as an informal guidance to quantify the amount of uncertainty involved in the value of a random variable or the outcome of a random process. (Shannon, 1948) In this research, entropy means the uncertainty and ambiguity associated with each feature. On the second-round interview, respondents were asked the following question:

Given the name of a feature (e.g., color combination, regularity, maintenance status), how much can you imagine a picture of the façade? By comparing the extent of imagination, we could estimate the entropy each feature has and select most informative ones among them.

Among 5 interviewees, 4 of them consider geometry distribution a vital feature in overall evaluation. In addition, all of them agree on the significance of color combination. Finally, we can easily remove redundant features from the list above by exclusion and narrow the list down to regularity and color combination. Additionally, all the features in the list above can be classified under “color” and “geometry” groups, which supports selecting regularity and color combination as a façade’s most significant features.

3.4 Quantitative research

A large-scale survey was conducted on multiple platforms worldwide to collect quantitative data on the façade evaluation criteria developed in Section 5.2. More than 120 people took part in it, and around 100 responses were valid during the process.

3.4.1 Mechanism of survey

Criteria:

The survey consists of four sections: (1) (a) regularity of the façade (b) likability of the regularity; (2) (a) color combination of the façade (b) likability of the color combination; (3) overall evaluation; (4) context. Each respondent was asked to rank 10 facades in the above categories. The process took approximately 8-12 mins for each respondent.

Likert score:

Various kinds of rating scales have been developed to measure attitudes directly. The most widely used is the Likert scale. It is applied to measure “attitudes, knowledge, perceptions, values, and behavioral changes, which involves a series of statements that respondents may choose from in order to rate their responses to evaluative questions” (Vogt, 1999). A Likert scale considers that an attitude has linear strength and intensity, which makes it measurable, for example, from dislike very much to like very much on a continuum.

Likert scales typically range from 2 to 10 values, with 3, 5, or, 7 being the most common. In this case, we use 6 values to exclude a neutral value (Deaton, 2010), since respondents sometimes tend to select the midpoint in online survey. Using an even number Likert scale ensures that respondents make a decision about each question; this adds credibility to the data.

	Please rank the façade's regularity	I like the regularity of this facade	Please rank the façade's color harmony	I like the color combination of this facade	Rank the aesthetic impression of this facade.	Context evaluation
Likert scale	1-6	1-6	1-6	1-6	1-6	1-6
Extreme definition - 1	Very low regularity	Strongly disagree	Very low color harmony	Strongly disagree	I like it very much	Very bad fit
Extreme definition - 6	Very high regularity	Strongly agree	Very high color harmony	Strongly agree	I dislike it very much	Very good fit

Table 3.4.1 Ranking criteria statement

The survey applied standard question formats for Likert Scales (Brown, 2010). For agreement measurement questions (likability), the extremes are “strongly disagree” and “strongly agree”; for value measurement questions (regularity, color harmony), they are “very low” and “very high”; for quality measurement (context), they are “very bad” and “very good”.

Learning effect:

In the online controlled experiment, the learning effect is the positive or negative effect of the intervention, which is only significant after a certain time. The effect may increase gradually over time, or there may be more dramatic level changes over time. (Analytics Toolkit, <https://www.analytics-toolkit.com/glossary/learning-effect/>)



Regularity



Color combination



Overall aesthetic impression



Context

*Figure 3.4.1 Ranking category of the same façade
Reprinted*

From(<https://www.sidewalklabs.com/>) used without permission

In this survey, the learning effect may cause a negative influence because the respondents will encounter pictures of the same façade several times for ranking based on different categories. Though these pictures are modified in Photoshop to highlight different features, people may still develop fixed attitudes towards a certain façade when going through the process. This can affect the ranking of the following features. In order to reduce the learning effect to some extent, we applied the “shuffle question order” function on online survey platforms. By randomizing the question sequence in each section. (Deaton, 2010)

Languages:

The survey was conducted using both Google Form and Questionnaire Star, in order to access subjects in China. The questions in Google Form are in English, while

Questionnaire Star has added a Chinese translation for each question for better accessibility. Considering the potential cultural and regional differences of the two groups, there might be a potential impact on the score of the façade. As a result, we use statistical methods to analyze the data of the two platforms to judge whether they are unified. Shown in the fig 4.3, there is no significant difference between the two groups, and they can be merged for further analysis.

3.4.2 Regularity

As is indicated in section 2.3, the current study focuses on judging the regularity of façades using mathematical (Janssen et al. 2013) and computational (Zhang et al. 2013) methods. We assume there is no vital difference between computational methods and manual methods in determining the regularity degree. We start by asking people to rank the regularity of facades using photos of them.

Regularity in a building facade means that the architectural elements (structure, cladding, windows, etc.) are arranged in a systematic way. For example, on a facade with high regularity the elements may have:

- Symmetry
- Duplicated patterns
- Edge alignment
- Concise geometry composition
- Regularity of material

On the other hand, a facade with very low regularity may not exhibit any alignment or patterns.



E.g. Façade with high regularity



E.g. Façade with low regularity (Sagrada Família by Gaudí)

Figure 3.4.2 Examples of facades' regularity

Reprinted from

(<https://zh.wikipedia.org/wiki/%E9%AB%98%E8%BF%AA%E7%9A%84%E5%BB%BA%E7%AD%91%E4%BD%9C%E5%93%81> and <https://www.pinterest.com/>) used without permission

Respondents were given the above instructions and examples to rank the regularity of certain facades. The photos were desaturated to exclude any potential color impact. In pre-test the survey provided the suggested score of facades in examples with a numerical example, but some respondents expressed some confusion about this. Therefore, in the final version of survey we asked respondents to rank the degree of regularity more qualitatively (e.g. a façade with high regularity) instead of labeling it with a specific score (a façade with 6 points) to avoid this confusion. People were then asked to rank how much they liked the regularity of given facades. This allowed us to study the relationship between regularity and likability of regularity. They were instructed to focus on the architectural elements of building and ignore distracting objects such as cars and trees in front of the building.

3.4.3 Color combination

Harmonious color combinations seem pleasing, while clashing color combinations seem unpleasant. The survey aims to find out how pleasing the color combinations on storefronts are to people.

Respondents were presented a series modified images of facades and asked to rate the color harmony of each facade and how much they like it. In this part of the survey, people did not need to explain or justify their choice. Often it can be difficult to say exactly what causes these impressions, other than that they feel "right" or "wrong".



E.g. facade with a clashing color combination



E.g. facade with a harmonious color combination

Figure 5.4.3 Examples of facades' color combination

Reprinted

from (<https://www.kooksaam.nl/worst-colors-to-paint-walls-kv.html> and <https://www.pinterest.com/olikstudio/restaurantes/>) used without permission

We applied the “Crystalize” function in Photoshop to get rid of unnecessary geometry details of façade photos so that respondents could focus on the color harmony evaluation. The degree of crystallization depends on the size of image. The values are manually tested to make sure the pixelation level in each photo is similar. Respondents also evaluated how much they like the façade’s color combination.

3.4.4 Overall evaluation

People rated their overall aesthetic preference for each facade based on their general impression. In order to see the intuitive choice of the respondent, we didn’t provide examples of what a good/ bad façade should look like in overall evaluation.

3.4.5 Context

For the context evaluation in the survey, the façade in question is marked with an obvious red arrow in the street-view photo for ranking in this section. Respondents need to rank how well the target façade fits with the nearby buildings. Some key points respondents are likely to consider in this evaluation are relative height, appropriateness of decorative elements, color harmony with nearby facades, etc. An example presented to respondents shows what a façade with a good fit in its context might look like.



Figure 3.4.5 Example of facade with a good fit in its context, Reprinted From(Mehdipour et al. 2020, Determining the Components Describing the Harmony-Contrast of the Color Combination in Residential Buildings Exterior, Int. J. Architect. Eng. Urban Plan, 31(1): 1-17 January 2021 DOI: 10.22068/ijaup.iust.2020.103) used without permission

3.5 Statistical analysis

Data processing: Due to the nature of a volunteer online survey, the data needs further processing to ensure the quality of analysis. Initial data processing dealt with identifying outliers, removing results from respondents who completed the survey more than once, and removing the results from respondents who have given contradictory or invalid

responses. For instance, some respondents ranked all the facades with 6 in every question, which is assumed to be an invalid response. Additionally, time spent on finishing the survey is another dimension for evaluation. It took respondents 7-10 minutes to fill out the survey in the pre-test. If the respondent in the final survey finished it within 2 minutes, it is considered to be an unqualified data point and will be removed.

Test description:

Name	Aim	Input	Output	Mechanism
Paired Wilcoxon test	Paired Wilcoxon test is used to test whether the paired data from diverse platforms are significantly different.	Paired data from the two platforms	P value	Functionally, it's identical to the paired t-test but used in nonparametric test
Cronbach Alpha analysis	To test whether collected data is valid or not	All data	Cronbach α	If a survey has n questions and the average correlation coefficient is r, then the standardized α coefficient of this survey is $\alpha=nr/[(n-1) r+1]$
Shapiro Wilk test	Test whether data under each question is distributed normally. It determines whether parametric tests should be used	Data categorized by each question	W and the P value	$W = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$
Spearman's correlation	a non-parametric test used to measure the strength of association between two variables.	Features such as regularity, color combination, etc.	Correlation table	$r_s = \rho_{R(X),R(Y)} = \frac{\text{cov}(R(X), R(Y))}{\sigma_{R(X)} \sigma_{R(Y)}}$
Nonparametric visualization analysis	Data visualization of distribution	Regularity & likability of regularity; color combination & likability of color combination, etc.	Graph	
Linear regression/curve regression	an approach for modelling the relationship between dependent and independent variables		Fitting line/curve	
Multiple linear regression	A statistical technique that uses several explanatory variables to	Dependent variables: regularity,	Fitting line	$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i$

predict the outcome of a response variable	color combination. Independent variables: overall evaluation score
--	--

Figure 3.5 Summary of statistical tests description, aim, data input & output and theoretical mechanism is listed.

4.Results

Based on the research proposal, we list the following hypothesis and aim to figure out their feasibility:

- (1) We can predict the likability of a façade using regularity and color harmony alone.
- (2) We can predict that if a façade is a good fit in context given the color and regularity features of buildings nearby.

For each hypothesis, there are several questions and sub-questions need figuring out. For example, regarding regularity, we planned to figure out not only whether there is a positive correlation between a façade’s regularity and its likability, but also where is the peak in mathematical model if any. In this case, we mainly focus on the first hypothesis because of insufficient facades samples to draw a meaningful conclusion on the second hypothesis: the built regression model didn’t pass the F test, which indicates the model is invalid mathematically. It requires further research with more facades samples to apply in survey.

4.1 Paired test towards multiple data sources

As is indicated in section 3.3.1, data collected from diverse survey platforms might impact the quality of the data and its analysis because of differences in how subjects respond to the surveys on different platforms. To determine this impact, it was necessary to perform a paired test to identify whether the data from the two platforms are compatible.

The Wilcoxon signed-rank test is a non-parametric statistical hypothesis test “used either to test the location of a population based on a sample of data, or to compare the locations of two populations using two matched samples” (Conover, 1999). For two matched samples, it is a paired difference test like the paired t-test in non-parametric approach.

For paired Wilcoxon test, the null hypothesis is there is no significant difference between the two groups. When p value is less than 0.05 or 0.01, the null hypothesis is rejected, which means the difference exists.

Paired Wilcoxon					
Items	Medium (P25, P75)		Mean difference (Paired1-Paired2)	z value	p
	Paired1	Paired2			
(1) Please rank the facade's regularity (1) 请对此立面的规律性打分 paired 1. Please rank the facade's regularity.	5.000(4.0,5.0)	4.500(4.0,5.0)	0.5	0.641	0.522
(2) Please rank the facade's regularity (2) 请对此立面的规律性打分 paired 2. Please rank the facade's regularity	3.500(2.0,4.0)	4.000(3.0,5.0)	-0.5	0.84	0.401
(4) Please rank the facade's regularity (4) 请对此立面的规律性打分 paired 4. Please rank the facade's regularity.	5.000(4.0,5.8)	4.000(3.0,5.0)	1	2.079	0.038*
(5) Please rank the facade's regularity (5) 请对此立面的规律性打分 paired 5. Please rank the facade's regularity.	4.000(4.0,4.8)	4.000(4.0,5.0)	0	0.436	0.663
(11) Please rank the facade's regularity (11) 请对此立面的规律性打分 paired 11. Please rank the facade's regularity.	3.000(2.0,4.0)	3.000(2.0,4.0)	0	0.698	0.485
(1) I like the regularity of this facade (1) 我喜欢此立面的规律性 paired 1. I like the regularity of this facade	4.000(4.0,5.0)	5.000(4.0,5.8)	-1	0.74	0.459
(2) I like the regularity of this facade (2) 我喜欢此立面的规律性 paired 2. I like the regularity of this facade	4.000(3.0,4.0)	4.000(3.0,5.0)	0	0.422	0.673
.....					
* p<0.05 ** p<0.01					

Table 4.1 Part of Table of Paired Wilcoxon, indicating the result of compare the paired data to identify whether there is significant difference between them. The pair with p value smaller than 0.05 indicates the difference exists.

According to Appendix C, in a total of 38 pairs of paired data, there are 2 pairs of paired data that show differences ($p < 0.05$): (1) For No.4 facade, there is a 0.05 level of significance ($p = 0.038 < 0.05$) in the regularity evaluation. (2) Another one is the overall aesthetic evaluation of No. 10 facade ($p = 0.019 < 0.05$)

Considering that 95% (37 of 39) of total number of pairs don't show significant differences (shown In Appendix C), and the difference between the two exception groups is not very large, we assume that there is no significant difference between the two groups of data as a whole, and they can be merged in the following analysis.

4.2 Reliability Test

Reliability Statistics (Cronbach α)			
Items	Corrected Item-Total Correlation (CITC)	Cronbach Alpha if Item Deleted	Cronbach α
(1) Please rank the facade's regularity	0.609	0.955	0.956
(2) Please rank the facade's regularity	0.705	0.954	
(4) Please rank the facade's regularity	0.569	0.955	
(5) Please rank the facade's regularity	0.601	0.955	
(11) Please rank the facade's regularity	0.678	0.954	
(11) I like the regularity of this façade	0.566	0.955	
(1) I like the regularity of this façade	0.586	0.955	
(2) I like the regularity of this façade	0.737	0.954	
(4) I like the regularity of this façade	0.675	0.954	
(5) I like the regularity of this façade	0.701	0.954	
(9) Please rank the facade's color harmony	0.458	0.955	
(1) Please rank the facade's color harmony	0.709	0.954	
(4) Please rank the facade's color harmony	0.68	0.954	
(5) Please rank the facade's color harmony	0.597	0.955	
(2) Please rank the facade's color harmony	0.603	0.955	
(4) I like the color combination of this façade	0.614	0.955	
(1) I like the color combination of this façade	0.676	0.954	
(5) I like the color combination of this façade	0.627	0.954	
(2) I like the color combination of this façade	0.619	0.955	
(9) I like the color combination of this façade	0.484	0.955	
(10) Rank your aesthetic impression of this façade	0.498	0.955	
(1) Rank your aesthetic impression of this façade	0.618	0.955	
(2) Rank your aesthetic impression of this façade	0.651	0.954	
(3) Rank your aesthetic impression of this façade	0.593	0.955	
(4) Rank your aesthetic impression of this façade	0.591	0.955	
(5) Rank your aesthetic impression of this façade	0.555	0.955	
(6) Rank your aesthetic impression of this façade	0.57	0.955	
(7) Rank your aesthetic impression of this façade	0.563	0.955	
(8) Rank your aesthetic impression of this façade	0.533	0.955	
(9) Rank your aesthetic impression of this façade	0.449	0.956	
(1) Rank the facade with the red arrow	0.542	0.955	
(2) Rank the facade with the red arrow	0.562	0.955	
(3) Rank the facade with the red arrow	0.666	0.954	
(4) Rank the facade with the red arrow	0.553	0.955	
(5) Rank the facade with the red arrow	0.645	0.954	
(6) Rank the facade with the red arrow	0.589	0.955	
(7) Rank the facade with the red arrow	0.403	0.956	
(8) Rank the facade with the red arrow	0.384	0.956	

Cronbach α (Standardized): 0.956

Table 4.2 Table of statistical tests description, Cronbach α score indicates the overall reliability score of data, “Cronbach Alpha if Item Deleted” shows the change of Cronbach α if the item is deleted

Cronbach’s alpha is a measure of internal consistency, that is, how closely related a set of items are as a group. It is considered to be a measure of scale reliability. Existing standards (Nunnally, 1967) consider value above 0.8 is great as the reliability of the test or scale; if the reliability coefficient is above 0.7, it is acceptable; otherwise, the scale needs to be redesigned.

According to the Table 4.2, the reliability coefficient value of the research data in this study is 0.956, higher than 0.9, which comprehensively indicates high reliability of the data quality and that it can be used for further analysis.

4.3 Normal distribution test

The quantitative research in this study uses Likert Scale in data collection, which is usually to be considered as an ordinal data. However, some (Jamieson, 2004) claim Likert Scale can also be treated as interval data in some cases. The difference between ordinal data and interval is that the intervals between the scale values are not equal for ordinal data. As a result, we performed a Shapiro Wilk test to identify whether the data is in normal distribution and can be analyzed with parametric methods or not.

The null hypothesis of Shapiro Wilk test is in normal distribution. In this case, if the p value is bigger than chosen alpha level (e.g., 0.05 or 0.01) we retain the null hypothesis and assume the data is in normal distribution; Otherwise, the null hypothesis is rejected, and data is assumed to be not normally distributed. (JMP, 2014)

We ran the test in the Google Collab Notebook (a cloud-based Python environment¹) with each facade by different features (regularity, color combination, etc.). As indicated in Figure 6.3, the P value for data acquired by each question is all below 0.05 (See Appendix D), showing none of them is normally distributed. In this case, nonparametric methods are preferred in the following procedure.

4.4 Correlation analysis

¹ https://colab.research.google.com/?utm_source=scs-index

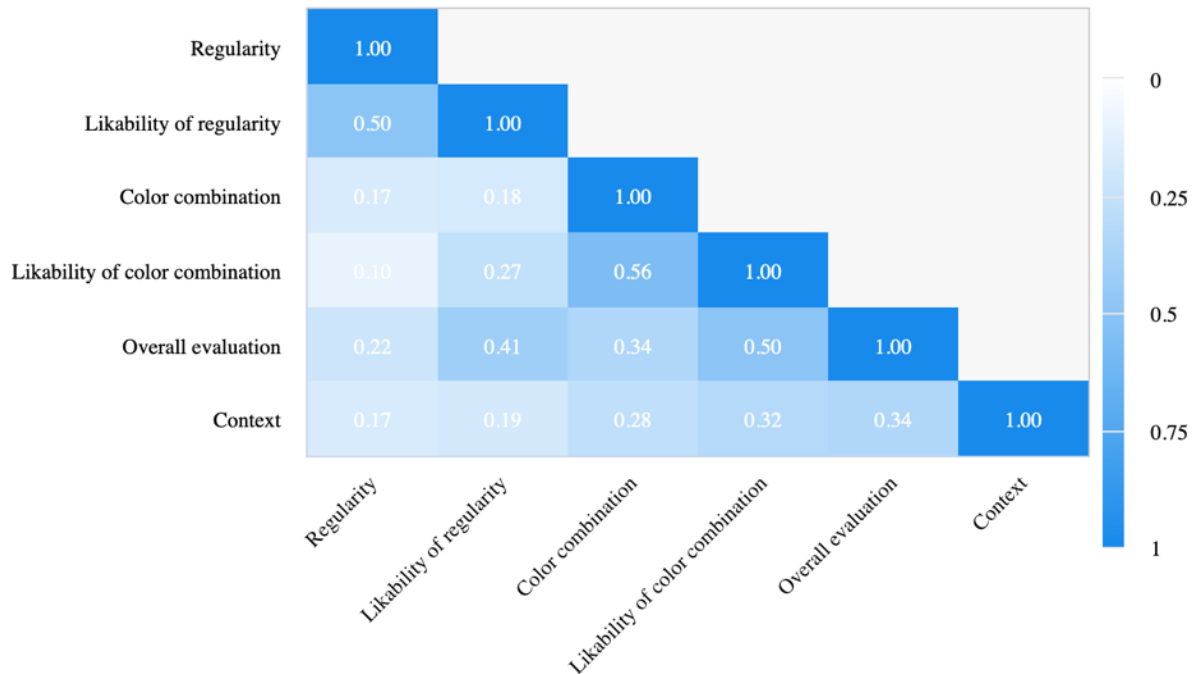


Table 4.4.1 Spearman correlation of regularity, likability of regularity, color combination, likability of color combination, overall evaluation and context. The gradient indicates the correlation extent of different variables.

Due to the nature of the data which doesn't follow a normal distribution (Chapter 3.1), a Spearman test was performed for correlation analysis. This is a nonparametric measure of rank correlation (statistical dependence between the rankings of two variables). It assesses how well the relationship between two variables can be described using a monotonic function.

The Spearman correlation between two variables is equivalent to the Pearson correlation between the rank values of those two variables; while Pearson's correlation assesses linear relationships, Spearman's correlation assesses monotonic relationships. Additionally, Spearman's coefficient is appropriate for both continuous and discrete ordinal variables. (Chorder et al. 2014)

The four features of Regularity, Likability of regularity, Color combination, Overall evaluation, and Context are significant. The respective correlation coefficient values are 0.50, 0.17, 0.22, 0.17 (illustrated in Table 4.4.11) Regularity and Likability of regularity, Color combination, Overall evaluation, and Context have a positive correlation between each other which is indicated by positive correlation coefficients.

The table also indicates the relatively strong relationship between Color combination and likability of color combination (0.56), overall evaluation and likability of color

combination (0.41) and overall evaluation and likability of regularity (0.50). We will further analyze these features and use them when formulating a mathematical model. At the same time, there is no significant relationship between Regularity and Likability of color combination, and the correlation coefficient value is close to 0(0.101). We also run the Pearson test as a comparison and find that most values are similar, with a 0.07 interval change at most. It may indicate some data is closed to normal distribution even though they don't pass the Shapiro Wilk test.

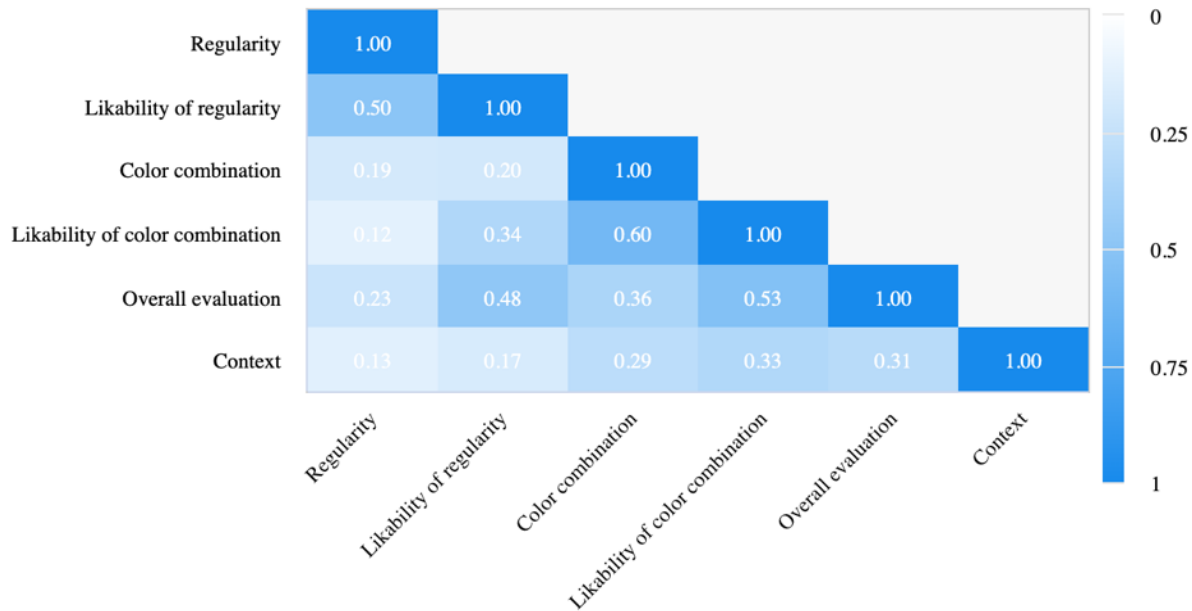


Table 4.4.2 Pearson correlation of regularity, likability of regularity, color combination, likability of color combination, overall evaluation, and context. The gradient indicates the correlation extent of different variables.

4.5 Relationship between regularity score and likability

Nonparametric test								
	Likability of regularity Median (P ₂₅ , P ₇₅)						Kruskal-	<i>p</i>
	1.0 (n=7)	2.0 (n=30)	3.0 (n=58)	4.0 (n=96)	5.0 (n=79)	6.0 (n=38)	Wallis H	
Regularity	1.000(1.0, 5.0)	3.000(2.0, 4.3)	4.000(3.0, 4.0)	4.000(4.0, 5.0)	5.000(4.0, 5.0)	6.000(5.0, 6.0)	80.597	0.000*

* *p*<0.05 ** *p*<0.01

Table 4.5.1 Nonparametric test (Kruskal-Wallis H) of regularity and likability

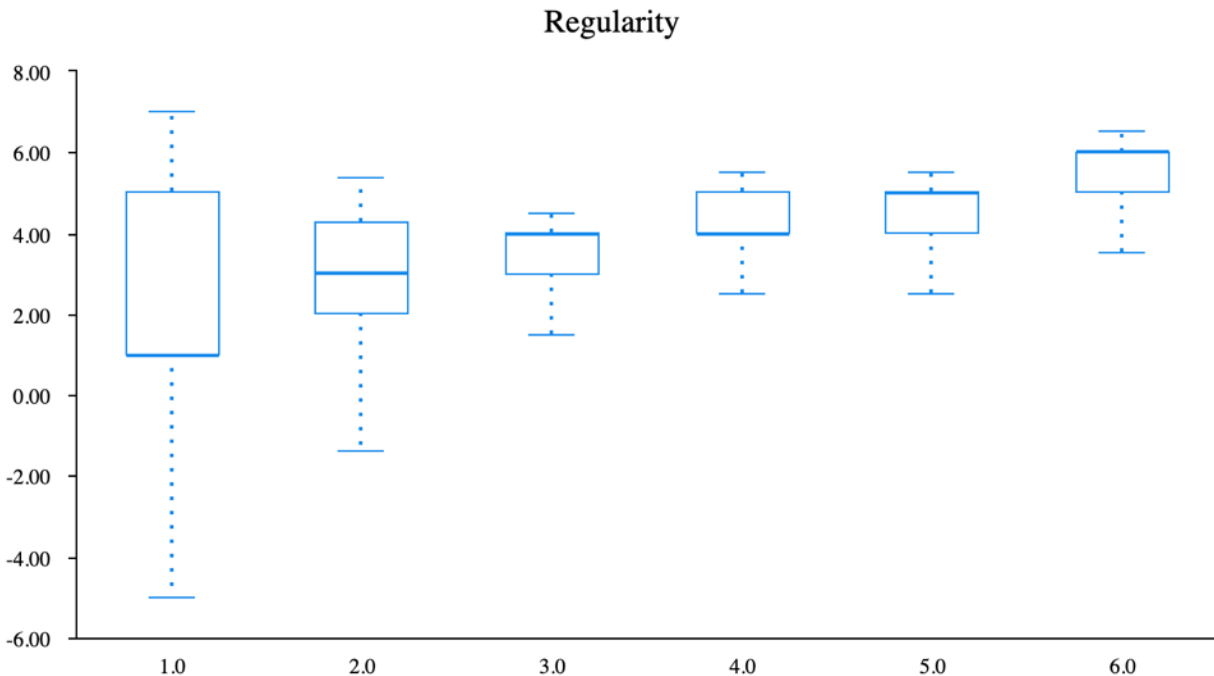


Figure 4.5.1 Visualization of data distribution (regularity & likability)

Using the SPSS (Statistical Package for the Social Sciences²) tools, we produced a table of relationships between regularity and likability. As can be seen from the above table (4.5.1). The null hypothesis is that there is no significant relationship between regularity and likability of regularity. All samples of different Likability of regularity showed significant ($p < 0.05$) for Regularity, which means that samples of different Likability of regularity had differences in Regularity.

By analyzing the median in Table 4.5.2, it is found that people's rankings on Likability are largely diverse when the regularity of facades are low (1.0 or 2.0). As the regularity increases, the likability score rises correspondently, which indicates a positive correlation.

In order to explore these relationships in more detail, we performed a curvilinear regression model by quadratic. We both build the linear and quadratic regression model and find that even quadratic regression model has a higher R^2 value, the difference is tiny (0.02). Besides the quadratic regression model looks like a linear way. We decide to display Linear regression model in the figure.

² <https://spssau.com/indexs.html>

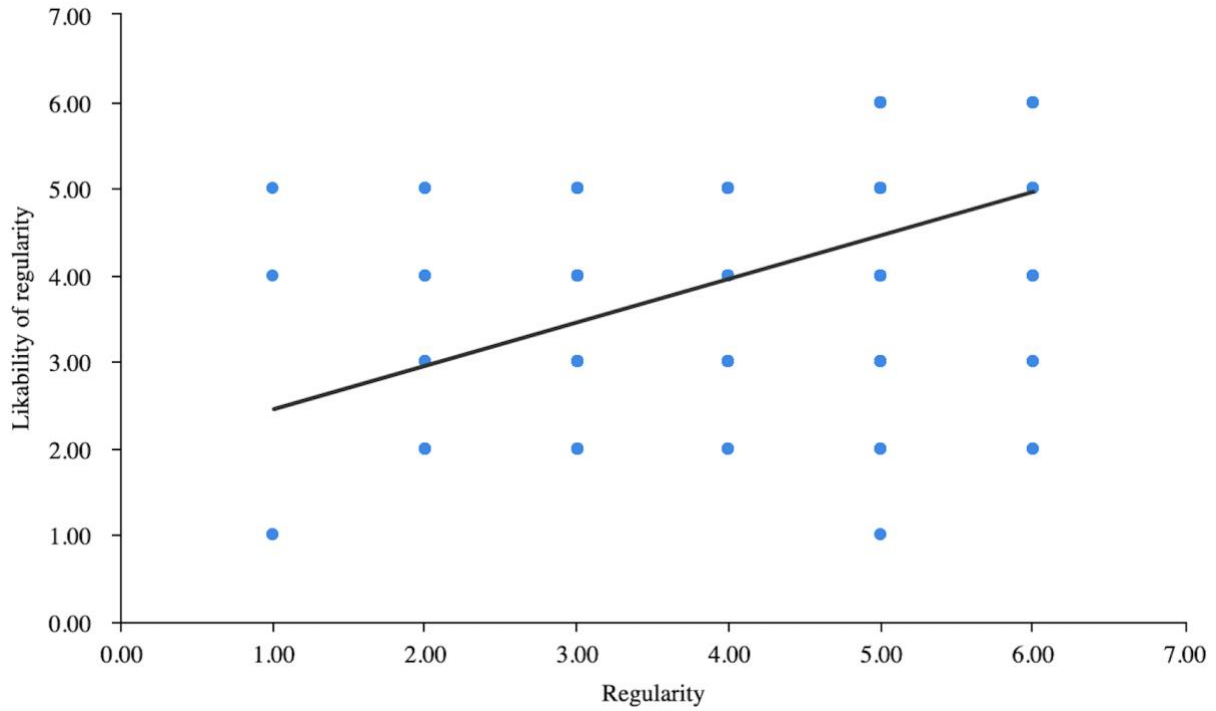


Figure 4.5.2 Linear regression (regularity & likability)

Model Summary

R^2	Adj. R^2	Std. Error of the Estimate	AIC	BIC	N
0.274	0.270	1.095	1165.286	1177.146	385

Table 4.5.2 Model Summary of curve regression (regularity & likability)

ANOVA

	Sum of Squares	df	Mean Square	F	p value
Regression	172.585	2	86.293	72.000	0.000**
Residual	457.830	382	1.199		
Total	630.416	384			

* $p < 0.05$ ** $p < 0.01$

Table 4.5.3 ANOVA test (regularity & likability)

Curvilinear regression refers to a method for regression analysis of variables with nonlinear relationships. It is a non-linear relationship in relational form, but it can be transformed into a linear relationship through various mathematical transformations,

and finally a regression model is established. It is used to study the influence relationship between X and Y, whether there is an influence relationship and the direction of influence.

According to Figure 4.5.2 and Table 4.5.2, here is the equation of regularity & likability:

$$f(x) = 1.783 + 0.557 * x - 0.004 * x^2$$

f(x): likability of regularity

x: regularity

In general, the two variables are in a positive correlation: as the regularity of the façade increases, so does people’s likability for the facade. At the same time, when the regularity is close to the highest level (6), the curvature of the curve suddenly decreases, which makes it appear to be flattened. The tendency corresponds with some prior research (Malewczyk, 2022), indicating facades with high regularity have more aesthetic preference from people.

Limitation: It should be noted that the curvilinear regression might lose preciseness to some extent when applied to ordinal data. Even though Some people maintain that while technically the Likert scale item is ordered, using it in parametric tests is valid in some situations. For example, Lubke & Muthen (2004) found that it is possible to find true parameter values in factor analysis with Likert scale data, if assumptions about skewness, number of categories, etc., were met. Likewise, Glass et al. (1972) found that F tests in ANOVA could return accurate p-values on Likert items under certain conditions. In this case, the ANOVA P value is less than 0.01, which means the model is approved and meaningful. However, according to R², regularity can only account for 27% of likability of regularity. There’s still space for exploration.

4.6 Relationship between color combination and likability

Nonparametric test								
	Likability of color combination Median (P ₂₅ , P ₇₅)						Kruska-	p
	1.0 (n=8)	2.0 (n=48)	3.0 (n=79)	4.0 (n=92)	5.0 (n=57)	6.0 (n=24)	Wallis	
							H	
Color combination	2.500(2.0, 3.0)	3.000(2.0, 4.0)	4.000(3.0, 4.0)	4.000(3.3, 5.0)	5.000(4.0, 5.0)	6.000(5.0, 6.0)	98.974	0.000**
	* p<0.05 ** p<0.01							

Table 4.6.1 Nonparametric test (Kruskal-Wallis H) of color combination and likability

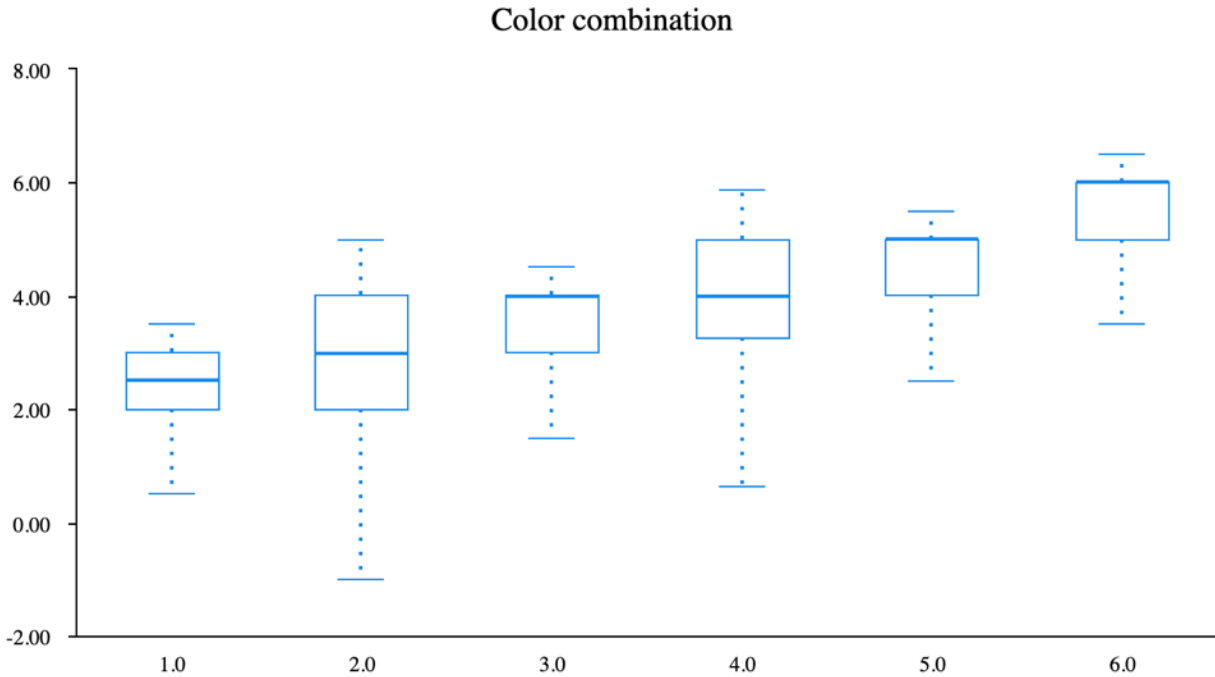


Figure 4.6.2 Visualization of data distribution (color combination & likability)

As is indicated in the table 4.6.2: The amount of likability of regularity exceeds two groups, so Kruskal-Wallis test statistic is used for analysis. All samples of different Likability of color combination showed significant ($p < 0.05$) for Color combination, which means that samples of different Likability of color combination had differences in Color combination. The correlation between color combination and likability of it exists. Additionally, people's ratings for Color combinations are more discrete. This allows the upper and lower bounds of the data to be stretched for a certain level, and the distance between the 25th and 75th percentile is longer (2, 4). It also indicates a positive correlation between color combination and likability of color combination.

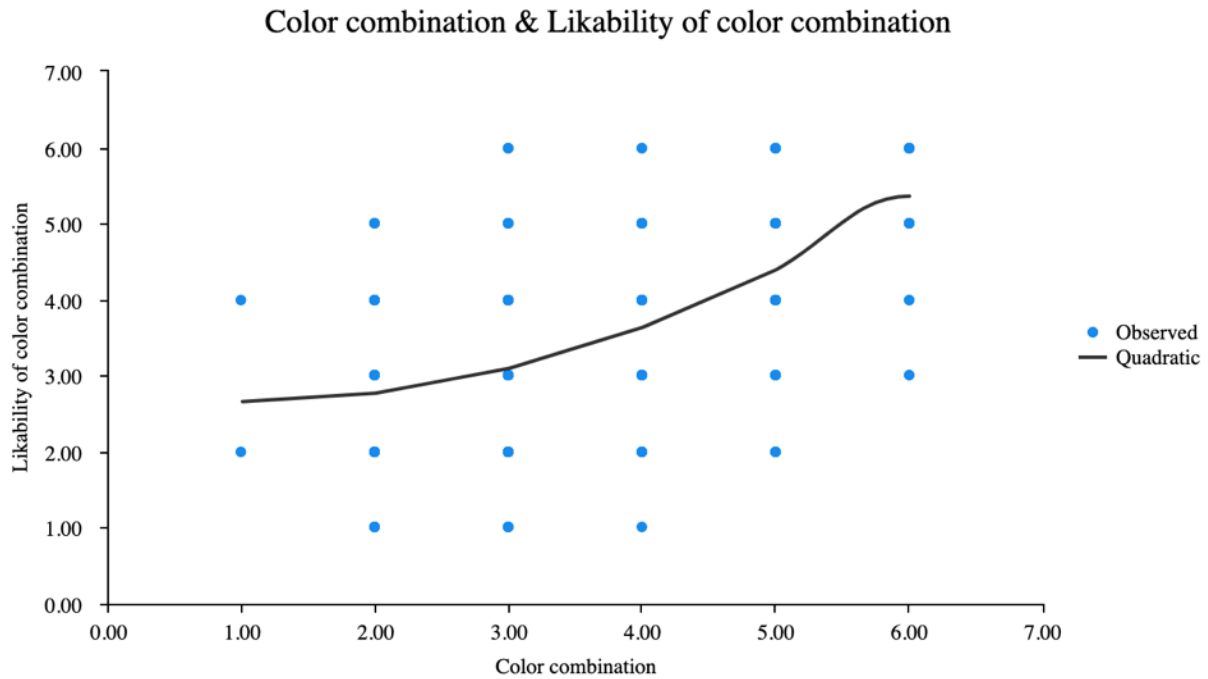


Figure 4.6.2 Curve regression (color combination & likability)

Model Summary

R^2	Adj. R^2	Std. Error of the Estimate	AIC	BIC	N
0.341	0.337	1.005	879.854	891.044	308

Table 4.6.2 Model Summary of curve regression (color combination & likability)

ANOVA

	Sum of Squares	df	Mean Square	F	p value
Regression	159.524	2	79.762	79.040	0.000**
Residual	307.788	305	1.009		
Total	467.312	307			

* $p < 0.05$ ** $p < 0.01$

Table 4.6.3 ANOVA test (combination & likability)

Based on the Figure 4.5.2 and Table 4.5.2, we can figure out the equation of color combination & likability:

$$f(x) = 2.762 - 0.211 * x + 0.107 * x^2$$

f(x): likability of color combination

x: r color combination

In general, the two variables are in a positive correlation: as the color combination of the façade increases, so does likability of it. Meanwhile, the curve is relatively flat in the 1-2 interval, and the difference in likability score is not large. When the degree of color combination rises, the slope of the curve increases, and the graph becomes steeper between 3-5.

Limitation: As is indicated in Chapter 6.5, using a regression method to analyze ordinal data might have potential problems in analysis because of the nature of data.

4.7 Relationship between overall score, regularity, and color combination

Multiple linear regression (MLR) is a statistical technique that models the linear relationship between the explanatory (independent) variables and response (dependent) variables. In order to figure out the relationship between overall evaluation and regularity & color combination at the same time, a MLR is performed through SPSS tools.

Parameter Estimates									
	Unstandardize	Standardize		t	p	VIF	R ²	Adj R ²	F
d	Coefficients								
	B	Std. Error	Beta						
Constant	1.648	0.261	-	6.327	0.000*	-	0.241	0.236	F (2,305)=48.297,p=0.000
Likability of regularity	0.351	0.049	0.367	7.226	0.000*	1.038			
Color combination	0.269	0.052	0.262	5.154	0.000*	1.038			

Dependent Variable: Overall evaluation

D-W: 1.693

* $p < 0.05$ ** $p < 0.01$

Table 4.7 Parameter of estimates of MLR

As is indicated in the Table 4.7, the model formula is:

$$f(x) = 1.648 + 0.351 * x_1 + 0.269 * x_2$$

$f(x)$ - Overall evaluation score

x_1 - Likability of regularity

x_2 - Color combination

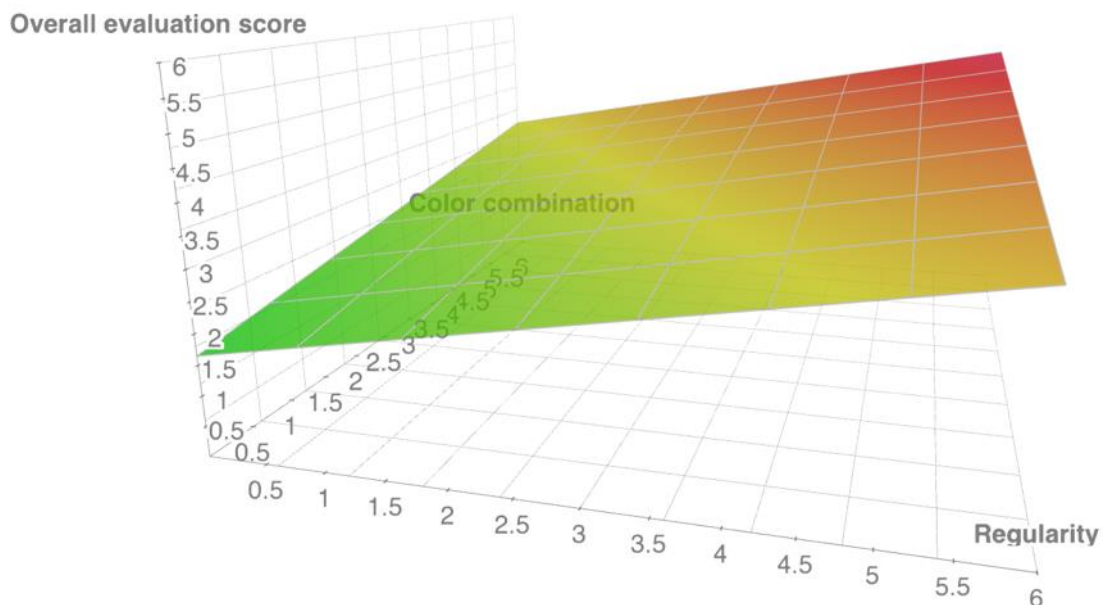


Figure 4.7 Relationship between overall evaluation score and regularity, color combination by 3D surface plot of the model formula (x-axis - regularity; y-axis - color combination; z-axis: - overall evaluation score)

The R-square value of the model is 0.241, which means that Likability of regularity and Color combination can account for 24.1% of the changes in Overall evaluation. When the F-test was performed on the model, it was found that the model passed the F-test ($F=48.297$, $p=0.000 < 0.05$), which means that at least one of Likability of regularity and Color combinations will have an impact on the Overall evaluation. The collinearity test shows that the VIF values in the model are all less than 5, which means that there is no collinearity problem; and the D-W value is near the number 2, which means that the model does not have autocorrelation, and there is no correlation between the sample data. Thus, the model is valid and here is the further analysis:

The regression coefficient value of Likability of regularity is 0.351 ($t=7.226$, $p=0.000<0.01$), which means that Likability of regularity will have a significant positive impact on Overall evaluation.

The regression coefficient value of Color combination is 0.269 ($t=5.154$, $p=0.000<0.01$), which means that Color combination will have a significant positive impact on Overall evaluation too.

In summary, Likability of regularity and Color combination both have a significant positive impact on Overall evaluation. However, it also indicates the given features can only account for about 25% for overall evaluation score based on low R^2 value, which demands further research of data processing and modeling. Although the low R-square value indicates they can only account for a limited content of the overall score, the high F score shows they are significant factors.

5. Discussion

- 1.The goal of this Thesis is not to replace the DRB. Instead, we aim to make their decision more consistent across projects and jurisdictions. We work on the fundamental work of building a framework to develop a system for evaluation of storefronts evaluation, on which few studies focus.
- 2.We use realistic photos of facades instead of drawings/ rendering pictures. In real world, DRB both conducts online review and on-site visits to have an impression of context integration.
- 3.We focus on daytime pictures of facades to exclude the impact of architectural lighting. Because the impact of architectural lighting on human comfort is another complex topic. (Pathak, et al. 2015)
- 4.Colors of glass are not counted in when we pick the primary color on the facades because of the complexities of reflection and refraction.
5. The reason we choose storefronts as the target building component is that storefronts are consistent in functionality, which makes them unified in styles and easier to extract features, compared to complex buildings like galleries or museums.
- 6.Due to the scale and preliminary nature of survey, we use traditional statistical method for analysis. In the future, if there is sufficient data, then we can use methods like machine learning to parameterize these models rather than our handcrafted statistical models.

6. Conclusion

The goal of Thesis is to explores the feasibility of leveraging visual features to quantify subjective criteria used by design review boards. It is a step towards filling the gap between designers and design review boards by integrating multiple subdivided

features into an overall study and applying it to architecture. In this case, we went through the process from background research, qualitative research, quantitative research, and statistical analysis to develop a mathematical model to predict the overall quality based on regularity and color combination. According to the obtained results, regularity and color combination appear to play an important role in people's overall evaluation of storefronts. These two features are proportional to likability individually as well: Facades with a high regularity are found to be more aesthetically pleasing than those with an irregular composition, so is color combination. However, the curve shape of the two-regression model is different: people's aesthetic preference for color combination is more diverse, while the correlation of regularity and likability of regularity is closer to a linear style, which means They are more sensitive to the change of regularity degree of facades compared to color factors. Based on features of regularity and color combination, we built a regression model to predict the overall quality of storefronts. This model predicts that people tend to have greater aesthetic preference for façades with high regularity and harmonious color combination. This model is not surprising or counter-intuitive and re-enforces our general understanding. It provides statistical support to existing architectural such as the leverage of concise and consist of geometry symbols(regularity) and organizing building with one primary color and 1-2 secondary colors (color combination). Additionally, concise facades with high regularity are more practical in storefronts design, which can be a helpful guideline to both DRB and architects.

Formulating general guidelines for DRB and architects on the basis of this study would be too premature. The conducted research should be treated as a starting point for further analyses, deepening the subject of the perception of building features and limited ranking scales in each question. First of all, attention should be paid to the problem of the scale of the regularity and color combination. Since we apply Likert Scale in quantitative research, there are only 6 degrees for respondents to rank, resulting in the difficulty of dealing with non-continuous variables. Supposing that people use a broader ranking scale for each question, we might be able to assume they are interval variables and apply more parametric methods of data analysis. Secondly the context part of the survey did not yield useful results. It needs further research to figure out the correlation of buildings to their contexts with given features like color combination and regularity.

In general, the research makes a start in system development for subjective evaluation of storefronts. The results of such analyses could allow for a more conscious design, as they would give designers and DRB an understanding of what reactions and perceptions are generated by specific aesthetic treatments of facades.

Reference

1. Christel Chamaret and Fabrice Urban, No-reference Harmony-guided Quality Assessment, CVPR2013
2. Malewczyk, M.; Taraszkiewicz, A.;Czyż ,P. Preferences of the Facade Composition in the Context of Its Regularity and Irregularity. *Buildings* 2022,12,169. <https://doi.org/10.3390/buildings12020169>
3. Christel Chamaret, Color harmony: experimental and computational modeling. Image Processing [eess.IV]. *Université Rennes 1*, 2016. English. NNT: 2016REN1S015. tel-01382750
4. Hannah and Hyunsoo Lee, Computational Color Design Process Towards Aesthetic Community Revitalization, *DESIGN TOOLS DEVELOPMENT - Volume 1 - eCAADe 36 | 677-686*
5. P. Janssen, P. Loh, A. Raonic, M. A. Schnabel (eds.), *Protocols, Flows and Glitches, Proceedings of the 22nd International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2017*, 581-591. © 2017, The Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.
6. Miller, R.J.: 2007, Another slant on the oblique effect in drawings and paintings, *Empirical*
7. *Studies of the Arts*, **25**, 41-61.
8. Sales Dias, M., Eloy, S., Carreiro, M., Proença, P., Moural, A., Pedro, T., Freitas, J., Vilar, E., D, J. and Azevedo, S.: 2014, Designing Better Spaces for People. Virtual reality and biometric sensing as tools to evaluate space use, *Proceedings of the 19th International Conference on CAADRIA 2014*, Kyoto, Japan, 739–748.
9. M. Mehdipour, S. A. Yazdanfar, A. Ekhlasi, B. Saleh Sedghpour, Determining the Components Describing the Harmony-Contrast of the Color Combination in Residential Buildings Exterior, *Int. J. Architect. Eng. Urban Plan*, 31(1): 1-17 January 2021, DOI: 10.22068/ijaup.iust.2020.103
10. Huang, C. (2018). Application and research of color in the façade design of elementary school buildings. Paper presented at *the 2nd International Conference on Art Studies: Science, Experience, Education, ICASSEE 2018*.
11. Li, K. R., Yang, Y. Q., & Zheng, Z. Q. (2020). Research on color harmony of building façades. *Color Research & Application*, 45(1), 105-119.
12. Jihyun Lee & Sungwon Kang & Chang-Ki Kim, Software architecture evaluation methods based on cost benefit analysis and quantitative decision making, *Empir Software, Eng* (2009) 14:453–475, DOI 10.1007/s10664-008-9094-4
13. Irena Gololov and Abraham Yezioro , A COMPUTER SYSTEM FOR MULTI-CRITERIA COMPARATIVE EVALUATION OF BUILDING ENVELOPES, *Proceedings: Building Simulation, 2007*
14. Hao Zhang, Kai Hu, Wei Jiang, Jinjie Lin, Daniel Cohen-Or Baoquan Chen, Layered analysis of irregular facades via symmetry maximization, *ACM Transactions on Graphics* Volume 32 Issue 4 July 2013 Article No.: 121pp 1–13 <https://doi.org/10.1145/2461912.2461923>

15. <https://www.seattle.gov/dpd/aboutus/news/events/designreview/upcomingreviews/>
16. <https://www.theurbanist.org/2021/09/22/fixing-seattles-design-review-process/>
17. <https://www.theurbanist.org/2021/07/19/its-time-to-overhaul-design-review/>
18. Kirk S. and Spreckelmeyer K. 1988. *Creative Design Decisions. A Systematic Approach to Problem Solving in Architecture*. New York.
19. Youdale J. 1983. *A comparative Study of Three Multiple Criteria Decision Methods to Ranking Decisions*. Simon Fraser University.
20. Zeng Jian, Xia Liangwa, *The Architecture Aesthetics*, China Architecture and Industry Press 2010
21. Zhang Dong, Fan Qindong, *The Art Performance of Architecture*, Art and Humanities, 2006
22. Hu Hua, *Analysis and Evaluation on Aesthetic Elements of Night Landscape*, Postgraduate Paper of Tianjin University, 2003
23. Angus Deaton, Instruments, Randomization, and Learning about Development, *Journal of Economic Literature* 48 (June 2010): 424–455 <http://www.aeaweb.org/articles.php?doi=10.1257/jel.48.2.424>
24. Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2001. The Colonial Origins of Comparative Development: An Empirical Investigation. *American Economic Review*, 91(5): 1369–1401.
25. Jamieson, S., 2004. Likert scales: how to (ab) use them. *Medical education*, 38 (12), 1217–1218.
26. Boone, H.N. and Boone, D.A., 2012. Analyzing likert data. *Journal of extension*, 50 (2), 1–5.
27. Kuzon Jr, W.M., Urbanchek, M.G., and McCabe, S., 1996. The seven deadly sins of statistical analysis. *Annals of plastic surgery*, 37 (3), 265–272.
28. Chigara, B., Moyo, T., & Mudzengerere, F. H. (2013). An Analysis of Cost Management Strategies Employed By Building Contractors on Projects in Zimbabwe. *International Journal of Sustainable Construction Engineering and Technology*, 4(2), 1–13. Retrieved from
29. Nur Najihah Abu Bakar, Mohammad Yusri Hassan, Hayati Abdullah, Hasimah Abdul Rahman, Md Pauzi Abdullah, Faridah Hussin, Masalah Bandi, Energy efficiency index as an indicator for measuring building energy performance: A review, *Renewable and Sustainable Energy Reviews*, Volume 44, 2015, Pages 1-11, ISSN 1364-0321, <https://doi.org/10.1016/j.rser.2014.12.018>.
30. <https://www.seattle.gov/sdci>
31. Nunnally, J. C. (1967). *Psychometric theory*. New York, NY: McGraw-Hill.
32. Shannon, Claude E. (July 1948). A Mathematical Theory of Communication. *Bell System Technical Journal*. 27 (3): 379–423. [doi:10.1002/j.1538-7305.1948.tb01338.x](https://doi.org/10.1002/j.1538-7305.1948.tb01338.x). [hdl:10338.dmlcz/101429](https://hdl.handle.net/10338.dmlcz/101429)

Appendix A: Questionnaire A

Facade Evaluation Survey

Welcome, and thank you for taking part in this study. Your participation will help us better understand how people perceive the quality of street facades. To explore this, we will ask you several questions to get your feedback. In total, this survey contains 6 parts and will take about 10 minutes. Your participation and candid feedback will be very helpful to us. Please take the survey freely, and feel free to quit the survey if you feel uncomfortable about it.

PS: This survey contains a completion code for SurveyStream.io.

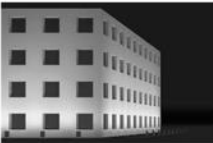
Required

Part 1: Regularity Evaluation

Regularity in building facade means that the architectural elements (windows, columns, moldings, etc.) are arranged in a systematic way. For example, in a facade with high regularity the elements may be distributed with aligned edges and in evenly determined intervals.

A facade with very low regularity may not exhibit any alignment or patterns. You will be presented a series of black and white images of facades and asked to rate their regularity and how much you like the regularity. Please base on the impression you have of the facade, not on the actual dimensions of the facade.

Example facade with low regularity



Example facade with low regularity



Part 1: Regularity evaluation

1. I like the regularity of this facade.*



Part 1: Regularity evaluation

1 2 3 4 5 6

Strongly disagree Strongly agree

2. I like the regularity of this facade.*



Part 1: Regularity evaluation

1 2 3 4 5 6

Strongly disagree Strongly agree

3. I like the regularity of this facade.*



Part 1: Regularity evaluation

1 2 3 4 5 6

Strongly disagree Strongly agree

4. I like the regularity of this facade.*



Part 1: Regularity evaluation

1 2 3 4 5 6

Strongly disagree Strongly agree

5. I like the regularity of this facade.*



Part 1: Regularity evaluation

1 2 3 4 5 6

Strongly disagree Strongly agree

6. Please rank the facade's regularity.*



Part 1: Regularity evaluation

1 2 3 4 5 6

Very low regularity Very high regularity

7. Please rank the facade's regularity.*



Part 1: Regularity evaluation

1 2 3 4 5 6

Very low regularity Very high regularity

8. Please rank the facade's regularity.*



Part 1: Regularity evaluation

1 2 3 4 5 6

Very low regularity Very high regularity

9. Please rank the facade's regularity.*



Part 1: Regularity evaluation

1 2 3 4 5 6

Very low regularity Very high regularity

10. Please rank the facade's regularity.*



Part 1: Regularity evaluation

1 2 3 4 5 6

Very low regularity Very high regularity

Part 2: Color coordination evaluation

Perceptive color coordination means planning which coloring color combinations mean emphasis. The survey aims to find out how planning the color combinations are distributed and to give you the opportunity to provide meaningful ratings of facades and asked to rate the color harmony of each facade and how much you like it.

In the next part of the survey you will be asked to evaluate or quality your choice. Please do not be difficult in any way, what counts these impressions, what does that mean "high" or "strong".

Example facade with a clashing color combination



Example facade with a harmonious color combination



Part 2: Color combination evaluation

11. Please rank the facade's color harmony *



Very low 1 2 3 4 5 6 Very high

12. I like the color combination of this facade. *



Strongly disagree 1 2 3 4 5 6 Strongly agree

13. Please rank the facade's color harmony *



Very low 1 2 3 4 5 6 Very high

14. Please rank the facade's color harmony *



Very low 1 2 3 4 5 6 Very high

15. Please rank the facade's color harmony *



Very low 1 2 3 4 5 6 Very high

16. Please rank the facade's color harmony *



Very low 1 2 3 4 5 6 Very high

17. I like the color combination of this facade. *



Strongly disagree 1 2 3 4 5 6 Strongly agree

18. I like the color combination of this facade. *



Strongly disagree 1 2 3 4 5 6 Strongly agree

19. I like the color combination of this facade. *



Strongly disagree 1 2 3 4 5 6 Strongly agree

20. I like the color combination of this facade. *



Strongly disagree 1 2 3 4 5 6 Strongly agree

Part 3: Overall evaluation
Please rate the overall quality of each facade based on your general impression. The higher the score, the more you like it.

21. 10. Rank your aesthetic impression of this facade. *



Rank only one item
1 2 3 4 5 6
I dislike it very much I like it very much

22. 1. Rank your aesthetic impression of this facade. *



Rank only one item
1 2 3 4 5 6
I dislike it very much I like it very much

23. 2. Rank your aesthetic impression of this facade. *



Rank only one item
1 2 3 4 5 6
I dislike it very much I like it very much

24. 3. Rank your aesthetic impression of this facade. *



Rank only one item
1 2 3 4 5 6
I dislike it very much I like it very much

25. 4. Rank your aesthetic impression of this facade. *



Rank only one item
1 2 3 4 5 6
I dislike it very much I like it very much

26. 5. Rank your aesthetic impression of this facade. *



Rank only one item
1 2 3 4 5 6
I dislike it very much I like it very much

27. 6. Rank your aesthetic impression of this facade. *



Rank only one item
1 2 3 4 5 6
I dislike it very much I like it very much

28. 7. Rank your aesthetic impression of this facade. *



Rank only one item
1 2 3 4 5 6
I dislike it very much I like it very much

29. 9. Rank your aesthetic impression of this facade. *



Rank only one item
1 2 3 4 5 6
I dislike it very much I like it very much

30. 8. Rank your aesthetic impression of this facade. *



Rank only one item
1 2 3 4 5 6
I dislike it very much I like it very much

Part 4: Context evaluation
Please rate how well each facade fits in its local context. Some buildings fit well in their context, others do not. A facade with a high score may have simple colors, height, materials, and glass ratio harmony with nearby facades. A facade with a low score will feel like it does not fit.

Example facade with a good fit in its context



Part 4: Context evaluation

31. 1. Rank the facade with the red arrow. *



Rank only one item
1 2 3 4 5 6
Very bad fit Very good fit

32. 2. Rank the facade with the red arrow. *



Rank only one item
1 2 3 4 5 6
Very bad fit Very good fit

33. 3. Rank the facade with the red arrow *



Mark only one box

1 2 3 4 5 6
Very bad fit Very good fit

34. 4. Rank the facade with the red arrow *



Mark only one box

1 2 3 4 5 6
Very bad fit Very good fit

35. 6. Rank the facade with the red arrow *



Mark only one box

1 2 3 4 5 6
Very bad fit Very good fit

35. 5. Rank the facade with the red arrow *



Mark only one box

1 2 3 4 5 6
Very bad fit Very good fit

37. 7. Rank the facade with the red arrow *



Mark only one box

1 2 3 4 5 6
Very bad fit Very good fit

38. 8. Rank the facade with the red arrow *



Mark only one box

1 2 3 4 5 6
Very bad fit Very good fit

End of survey. Thank you for your participation!

Appendix B: Questionnaire B

Facade Evaluation Survey

Welcome, and thank you for taking part in this study. Your participation will help us better understand how people evaluate the quality of storefront facades. To explore this, we will ask you some questions to get your feedback. In total, this survey contains 8 parts and will take about 8 minutes. Your participation and candid feedback will be very helpful to us. Please take the survey freely, and feel free to quit the survey if you feel uncomfortable about it.

P.S.: This survey contains a completion code for SurveySav®.

* Required

Part 1: Regularity Evaluation

Regularity in a building facade means that the architectural elements (windows, cladding, awnings, etc.) are arranged in a systematic way. For example, on a facade with high regularity the elements may be distributed with equal spacing, and in a clearly discernible pattern.

A facade with very low regularity may not exhibit any alignment or pattern.

You will be presented a certain black and white image of a facade and asked to rate their regularity and how much you like the regularity. Please focus on the architectural elements of building, and ignore distracting objects such as cars and trees in front of building.

Example: facade with high regularity



Example: facade with low regularity



Part 1: Regularity Evaluation

1. I like the regularity of this facade. *



Mark only one circle.

1 2 3 4 5 6
Strongly disagree Strongly agree

2. I like the regularity of this facade. *



Mark only one circle.

1 2 3 4 5 6
Strongly disagree Strongly agree

3. I like the regularity of this facade. *



Mark only one circle.

1 2 3 4 5 6
Strongly disagree Strongly agree

4. I like the regularity of this facade. *



Mark only one circle.

1 2 3 4 5 6
Strongly disagree Strongly agree

5. I like the regularity of this facade. *



Mark only one circle.

1 2 3 4 5 6
Strongly disagree Strongly agree

10. I like the regularity of this facade. *



Mark only one circle.

1 2 3 4 5 6
Strongly disagree Strongly agree

7. I like the regularity of this facade. *



Mark only one circle.

1 2 3 4 5 6
Strongly disagree Strongly agree

8. I like the regularity of this facade. *



Mark only one circle.

1 2 3 4 5 6
Strongly disagree Strongly agree

9. I like the regularity of this facade. *



Mark only one circle.

1 2 3 4 5 6
Strongly disagree Strongly agree

6. I like the regularity of this facade. *



Mark only one circle.

1 2 3 4 5 6
Strongly disagree Strongly agree

11. 2. I like the regularity of this facade. *



Mark only one oval.
1 2 3 4 5 6
Strongly disagree Strongly agree

12. 1. Please rank the facade's regularity. *



Mark only one oval.
1 2 3 4 5 6
Very low regularity Very high regularity

13. 2. Please rank the facade's regularity. *



Mark only one oval.
1 2 3 4 5 6
Very low regularity Very high regularity

14. 4. Please rank the facade's regularity. *



Mark only one oval.
1 2 3 4 5 6
Very low regularity Very high regularity

15. 5. Please rank the facade's regularity. *



Mark only one oval.
1 2 3 4 5 6
Very low regularity Very high regularity

16. 11. Please rank the facade's regularity. *



Mark only one oval.
1 2 3 4 5 6
Very low regularity Very high regularity

17. 3. Please rank the facade's regularity. *



Mark only one oval.
1 2 3 4 5 6
Very low regularity Very high regularity

18. 6. Please rank the facade's regularity. *



Mark only one oval.
1 2 3 4 5 6
Very low regularity Very high regularity

19. 7. Please rank the facade's regularity. *



Mark only one oval.
1 2 3 4 5 6
Very low regularity Very high regularity

20. 9. Please rank the facade's regularity. *



Mark only one oval.
1 2 3 4 5 6
Very low regularity Very high regularity

21. 10. Please rank the facade's regularity. *



Mark only one oval.
1 2 3 4 5 6
Very low regularity Very high regularity

22. 12. Please rank the facade's regularity. *



Mark only one oval.
1 2 3 4 5 6
Very low regularity Very high regularity

23. 8. Please rank the facade's regularity *



Mark only one oval

1 2 3 4 5 6

Very low regularity Very high regularity

Part 2: Color combination evaluation

harmonious color combinations seem pleasing, while clashing color combinations seem unpleasant. The survey aims to find out how pleasing the color combinations on storefronts are to you. You will be presented a series of real-life images of facades and asked to rate the color harmony of each facade and how much you like it. In this part of the survey, you do not need to explain or justify your choice. Often it can be difficult to say exactly what causes these impressions, other than that they feel "right" or "wrong".

Example: facade with a clashing color combination



Example: facade with a harmonious color combination



Part 2: Color combination evaluation

24. 9. Please rank the facade's color harmony *



Mark only one oval

1 2 3 4 5 6

Very low Very high

25. 3. Please rank the facade's color harmony *



Mark only one oval

1 2 3 4 5 6

Very low Very high

26. 6. Please rank the facade's color harmony *



Mark only one oval

1 2 3 4 5 6

Very low Very high

27. 7. Please rank the facade's color harmony *



Mark only one oval

1 2 3 4 5 6

Very low Very high

28. 10. Please rank the facade's color harmony *



Mark only one oval

1 2 3 4 5 6

Very low Very high

29. 12. Please rank the facade's color harmony *



Mark only one oval

1 2 3 4 5 6

Very low Very high

30. 8. Please rank the facade's color harmony *



Mark only one oval

1 2 3 4 5 6

Very low Very high

31. 3. I like the color combination of this facade. *



Mark only one oval

1 2 3 4 5 6

Strongly disagree strongly agree

32. 6. I like the color combination of this facade. *



Mark only one oval

1 2 3 4 5 6

Strongly disagree strongly agree

33. 7. I like the color combination of this facade. *



Mark only one oval.

1 2 3 4 5 6

Strongly disagree Strongly agree

34. 9. I like the color combination of this facade. *



Mark only one oval.

1 2 3 4 5 6

Strongly disagree Strongly agree

35. 10. I like the color combination of this facade. *



Mark only one oval.

1 2 3 4 5 6

Strongly disagree Strongly agree

36. 12. I like the color combination of this facade. *



Mark only one oval.

1 2 3 4 5 6

Strongly disagree Strongly agree

37. 8. I like the color combination of this facade. *



Mark only one oval.

1 2 3 4 5 6

Strongly disagree Strongly agree

38. 1. Please rank the facade's color harmony. *



Mark only one oval.

1 2 3 4 5 6

Very low Very high

39. 4. Please rank the facade's color harmony. *



Mark only one oval.

1 2 3 4 5 6

Very low Very high

40. 5. Please rank the facade's color harmony. *



Mark only one oval.

1 2 3 4 5 6

Very low Very high

41. 2. Please rank the facade's color harmony. *



Mark only one oval.

1 2 3 4 5 6

Very low Very high

42. 4. I like the color combination of this facade. *



Mark only one oval.

1 2 3 4 5 6

Strongly disagree Strongly agree

43. 5. I like the color combination of this facade. *



Mark only one oval.

1 2 3 4 5 6

Strongly disagree Strongly agree

44. 2. I like the color combination of this facade. *



Mark only one oval.

1 2 3 4 5 6

Strongly disagree Strongly agree

45. 1. I like the color combination of this facade. *



Mark only one oval.

1 2 3 4 5 6
Strongly disagree Strongly agree

Part 3:
Consent
Evaluation

Please rate the overall quality of each facade based on your general impression. The higher that score, the more you like it.

46. 10. Rank your aesthetic impression of this facade. *



Mark only one oval.

1 2 3 4 5 6
I dislike it very much I like it very much

47. 1. Rank your aesthetic impression of this facade. *



Mark only one oval.

1 2 3 4 5 6
I dislike it very much I like it very much

48. 2. Rank your aesthetic impression of this facade. *



Mark only one oval.

1 2 3 4 5 6
I dislike it very much I like it very much

49. 3. Rank your aesthetic impression of this facade. *



Mark only one oval.

1 2 3 4 5 6
I dislike it very much I like it very much

50. 4. Rank your aesthetic impression of this facade. *



Mark only one oval.

1 2 3 4 5 6
I dislike it very much I like it very much

51. 5. Rank your aesthetic impression of this facade. *



Mark only one oval.

1 2 3 4 5 6
I dislike it very much I like it very much

52. 6. Rank your aesthetic impression of this facade. *



Mark only one oval.

1 2 3 4 5 6
I dislike it very much I like it very much

53. 7. Rank your aesthetic impression of this facade. *



Mark only one oval.

1 2 3 4 5 6
I dislike it very much I like it very much

54. 9. Rank your aesthetic impression of this facade. *



Mark only one oval.

1 2 3 4 5 6
I dislike it very much I like it very much

55. 12. Rank your aesthetic impression of this facade. *



Mark only one oval.

1 2 3 4 5 6
I dislike it very much I like it very much

56. 8. Rank your aesthetic impression of this facade. *



Mark only one oval.

1 2 3 4 5 6
I dislike it very much I like it very much

Part 4:
Consent
Evaluation

Please rate how well each facade fits in its local context. Some buildings fit well in their context, others do not. A facade with a high score may have similar shapes, height, decorative, and good color harmony with nearby facades. A facade with a low score will feel like it does not fit.

Example: facade with a good fit in its context



Part 4: Context evaluation

57. 1. Rank the facade with the red arrow *



Very bad fit 1 2 3 4 5 6 Very good fit

58. 12. Rank the facade with the red arrow *



Very bad fit 1 2 3 4 5 6 Very good fit

59. 2. Rank the facade with the red arrow *



Very bad fit 1 2 3 4 5 6 Very good fit

60. 3. Rank the facade with the red arrow *



Very bad fit 1 2 3 4 5 6 Very good fit

61. 4. Rank the facade with the red arrow *



Very bad fit 1 2 3 4 5 6 Very good fit

62. 5. Rank the facade with the red arrow *



Very bad fit 1 2 3 4 5 6 Very good fit

63. 6. Rank the facade with the red arrow *



Very bad fit 1 2 3 4 5 6 Very good fit

64. 7. Rank the facade with the red arrow *



Very bad fit 1 2 3 4 5 6 Very good fit

65. 8. Rank the facade with the red arrow *



Very bad fit 1 2 3 4 5 6 Very good fit

Appendix C: Paired Wilcoxon Test

The appendix indicates the indicating the result of compare the paired data to identify whether there is significant difference between them. The pair with p value smaller than 0.05 indicates the difference exists.

Items	Medium (P25, P75)		Mean difference (Paired1 - Paired2)	z value	p
	Paired1	Paired2			
(1) Please rank the facade's regularity (1) 请对此立面的规律性打分 paired 1. Please rank the facade's regularity.	5.000(4.0,5.0)	4.500(4.0,5.0)	0.5	0.641	0.522
(2) Please rank the facade's regularity (2) 请对此立面的规律性打分 paired 2. Please rank the facade's regularity	3.500(2.0,4.0)	4.000(3.0,5.0)	-0.5	0.84	0.401
(4) Please rank the facade's regularity (4) 请对此立面的规律性打分 paired 4. Please rank the facade's regularity.	5.000(4.0,5.8)	4.000(3.0,5.0)	1	2.079	0.038*
(5) Please rank the facade's regularity (5) 请对此立面的规律性打分 paired 5. Please rank the facade's regularity.	4.000(4.0,4.8)	4.000(4.0,5.0)	0	0.436	0.663
(11) Please rank the facade's regularity (11) 请对此立面的规律性打分 paired 11. Please rank the facade's regularity.	3.000(2.0,4.0)	3.000(2.0,4.0)	0	0.698	0.485
(1) I like the regularity of this facade (1) 我喜欢此立面的规律性 paired 1. I like the regularity of this facade	4.000(4.0,5.0)	5.000(4.0,5.8)	-1	0.74	0.459
(2) I like the regularity of this facade (2) 我喜欢此立面的规律性 paired 2. I like the regularity of this facade	4.000(3.0,4.0)	4.000(3.0,5.0)	0	0.422	0.673
(4) I like the regularity of this facade (4) 我喜欢此立面的规律性 paired 4. I like the regularity of this facade	4.000(3.3,5.0)	5.000(4.0,5.0)	-1	0.915	0.36
(5) I like the regularity of this facade (5) 我喜欢此立面的规律性 paired 5. I like the regularity of this facade	4.000(4.0,5.0)	4.000(4.0,5.0)	0	0.949	0.343
(11) I like the regularity of this facade (11) 我喜欢此立面的规律性 paired 11. I like the regularity of this facade	4.000(3.0,4.8)	3.000(2.0,4.0)	1	1.026	0.305
(1) Please rank the facade's color harmony (1) 请对此立面的色彩和谐度评分 paired 1. Please rank the facade's color harmony	3.000(3.0,4.8)	4.000(3.3,5.0)	-1	1.311	0.19
(2) Please rank the facade's color harmony (2) 请对此立面的色彩和谐度评分 paired 2. Please rank the facade's color harmony	4.000(3.3,5.0)	4.000(4.0,5.0)	0	0.37	0.711

(4) Please rank the facade's color harmony (4) 请对此立面的色彩和谐度评分 paired 4. Please rank the facade's color harmony	4.000(3.0,5.0)	4.000(4.0,5.0)	0	1.14	0.253	2
(5) Please rank the facade's color harmony (5) 请对此立面的色彩和谐度评分 paired 5. Please rank the facade's color harmony	4.000(2.3,5.0)	4.000(3.0,5.0)	0	0.49	0.623	2
(9) Please rank the facade's color harmony (9) 请对此立面的色彩和谐度评分 paired 9. Please rank the facade's color harmony	4.500(4.0,5.0)	5.000(4.0,6.0)	-0.5	1.08	0.279	3
(1) I like the color combination of this facade (1) 我喜欢此立面的配色 paired 1. I like the color combination of this facade.	3.000(2.0,4.0)	4.000(3.0,4.0)	-1	0.65	0.516	
(2) I like the color combination of this facade (2) 我喜欢此立面的配色 paired 2. I like the color combination of this facade.	4.000(3.3,5.0)	4.000(3.0,5.0)	0	0.77	0.44	3
(4) I like the color combination of this facade (4) 我喜欢此立面的配色 paired 4. I like the color combination of this facade.	4.000(3.0,5.0)	4.000(3.0,5.0)	0	0.56	0.57	8
(5) I like the color combination of this facade (5) 我喜欢此立面的配色 paired 5. I like the color combination of this facade.	3.000(3.0,4.8)	3.500(3.0,4.8)	-0.5	0.38	0.7	5
(9) I like the color combination of this facade (9) 我喜欢此立面的配色 paired 9. I like the color combination of this facade.	5.000(3.0,6.0)	5.000(3.0,6.0)	0	0.26	0.794	1
(1) Rank your aesthetic impression of this facade (1) 请对这个立面的美学印象进行打分 paired 1. Rank your aesthetic impression of this facade.	4.000(3.0,4.0)	3.000(3.0,4.0)	1	1.09	0.275	1
(2) Rank your aesthetic impression of this facade (2) 请对这个立面的美学印象进行打分 paired 2. Rank your aesthetic impression of this facade.	4.000(3.0,4.8)	4.000(3.0,5.0)	0	0	1	
(3) Rank your aesthetic impression of this facade (3) 请对这个立面的美学印象进行打分 paired 3. Rank your aesthetic impression of this facade.	4.000(4.0,5.0)	4.500(3.0,5.0)	-0.5	0.21	0.834	
(4) Rank your aesthetic impression of this facade (4) 请对这个立面的美学印象进行打分 paired 4. Rank your aesthetic impression of this facade.	4.500(4.0,5.0)	4.000(3.0,5.0)	0.5	0.88	0.379	
(5) Rank your aesthetic impression of this facade (5) 请对这个立面的美学印象进行打分 paired 5. Rank your aesthetic impression of this facade.	4.000(4.0,5.0)	4.000(3.0,5.0)	0	0.94	0.346	2
(6) Rank your aesthetic impression of this facade (6) 请对这个立面的美学印象进行打分 paired 6. Rank your aesthetic impression of this facade.	3.500(2.0,4.0)	3.000(2.0,3.0)	0.5	1.32	0.186	3
(7) Rank your aesthetic impression of this facade (7) 请对这个立面的美学印象进行打分 paired 7. Rank your aesthetic impression of this facade.	4.000(3.0,5.0)	3.000(3.0,4.0)	1	1.43	0.152	3

(8) Rank your aesthetic impression of this facade 请对这个立面的美学印象进行打分 paired 8. Rank your aesthetic impression of this facade.	3.000(2.0,4.0)	3.000(2.0,4.0)	0	0.13 3	0.894
(9) Rank your aesthetic impression of this facade 请对这个立面的美学印象进行打分 paired 9. Rank your aesthetic impression of this facade.	4.000(4.0,5.0)	4.000(3.3,5.0)	0	0.48 2	0.63
(10) Rank your aesthetic impression of this facade (10) 请对这个立面的美学印象进行打分 paired 10. Rank your aesthetic impression of this facade.	4.000(4.0,5.0)	3.000(3.0,4.0)	1	2.34 6	0.019 *
(1) Rank the facade with the red arrow 色箭头的立面 paired 1. Rank the facade with the red arrow	3.500(3.0,4.8)	3.000(2.0,4.0)	0.5	1.18 8	0.235
(2) Rank the facade with the red arrow 色箭头的立面 paired 2. Rank the facade with the red arrow	4.000(3.0,5.0)	3.500(3.0,5.0)	0.5	0.07 9	0.937
(3) Rank the facade with the red arrow 色箭头的立面 paired 3. Rank the facade with the red arrow	4.000(3.0,4.0)	4.000(3.0,5.0)	0	0.02 9	0.977
(4) Rank the facade with the red arrow 色箭头的立面 paired 4. Rank the facade with the red arrow	5.000(5.0,6.0)	5.000(5.0,5.8)	0	0.11 4	0.909
(5) Rank the facade with the red arrow 色箭头的立面 paired 5. Rank the facade with the red arrow	5.000(4.0,5.0)	5.000(3.3,6.0)	0	0.69 3	0.488
(6) Rank the facade with the red arrow 色箭头的立面 paired 6. Rank the facade with the red arrow	4.000(3.0,5.0)	4.000(3.0,4.0)	0	0.87 8	0.38
(7) Rank the facade with the red arrow 色箭头的立面 paired 7. Rank the facade with the red arrow	6.000(5.0,6.0)	5.000(4.0,6.0)	1	1.62 7	0.104
(8) Rank the facade with the red arrow 色箭头的立面 paired 8. Rank the facade with the red arrow	5.500(4.0,6.0)	5.000(4.0,6.0)	0.5	0.92 8	0.354

* $p < 0.05$ ** $p < 0.01$

Appendix D: Result of Shapiro Wilk test

The table indicates the result of running Shapiro Wilk test to data by each question in the survey. The data is considered not to normally distributed if p-value is lower than 0.05. The conclusion is summarized to provide better readability and understanding.

Item number	W, p-value	Conclusion
1	(0.8957619071006775, 7.222034582810011e-06)	likely not a normal distribution
2	(0.9331409931182861, 0.0003851205692626536)	likely not a normal distribution
3	(0.86269611120224, 3.883979218244349e-07)	likely not a normal distribution
4	(0.8796542286872864, 1.6431730500698905e-06)	likely not a normal distribution
5	(0.9283199906349182, 0.000219102599658072)	likely not a normal distribution
6	(0.9313027858734131, 0.00030996158602647483)	likely not a normal distribution
7	(0.9158157706260681, 5.4753039876231924e-05)	likely not a normal distribution
8	(0.923782229423523, 0.0001308512728428468)	likely not a normal distribution
9	(0.9198021292686462, 8.423403778579086e-05)	likely not a normal distribution
10	(0.9140739440917969, 4.550155426841229e-05)	likely not a normal distribution
11	(0.8715928792953491, 8.16527574443171e-07)	likely not a normal distribution
12	(0.913252592086792, 4.172589615336619e-05)	likely not a normal distribution
13	(0.9141790866851807, 4.6010405640117824e-05)	likely not a normal distribution
14	(0.920299768447876, 8.895168866729364e-05)	likely not a normal distribution
15	(0.924822211265564, 0.00014707277296110988)	likely not a normal distribution
16	(0.9289259910583496, 0.0002349772403249517)	likely not a normal distribution
17	(0.9199661016464233, 8.575864194426686e-05)	likely not a normal distribution
18	(0.9317790865898132, 0.0003278179792687297)	likely not a normal distribution
19	(0.9220160841941833, 0.00010747962369350716)	likely not a normal distribution
20	(0.8345902562141418, 4.4113612318597006e-08)	likely not a normal distribution
21	(0.9221266508102417, 0.00010880499030463398)	likely not a normal distribution
22	(0.9239912033081055, 0.0001339529553661123)	likely not a normal distribution
23	(0.9116376638412476, 3.523288614815101e-05)	likely not a normal distribution
24	(0.9117833971977234, 3.5772332921624184e-05)	likely not a normal distribution
25	(0.9021573662757874, 1.3454104191623628e-05)	likely not a normal distribution
26	(0.8969284892082214, 8.077423444774467e-06)	likely not a normal distribution
27	(0.9140380024909973, 4.532913226285018e-05)	likely not a normal distribution
28	(0.923979640007019, 0.000133779103634879)	likely not a normal distribution
29	(0.9250184297561646, 0.00015036339755170047)	likely not a normal distribution
30	(0.8790546655654907, 1.5584807897539577e-06)	likely not a normal distribution
31	(0.9251729249954224, 0.000153010492795147)	likely not a normal distribution
32	(0.9283713102340698, 0.00022040218755137175)	likely not a normal distribution
33	(0.9262157678604126, 0.0001722073502605781)	likely not a normal distribution
34	(0.8404531478881836, 6.8096269956186e-08)	likely not a normal distribution

35	(0.877021312713623, 1.303910380556772e-06)	likely not a normal distribution
36	(0.9216358065605164, 0.00010304968600394204)	likely not a normal distribution
37	(0.8497234582901001, 1.3805809828681959e-07)	likely not a normal distribution
38	(0.8303356170654297, 3.238166712549173e-08)	likely not a normal distribution
39	(0.881818413734436, 3.4983073419425637e-05)	likely not a normal distribution
40	(0.9331809878349304, 0.0029936572536826134)	likely not a normal distribution
41	(0.8524950742721558, 4.220693426759681e-06)	likely not a normal distribution
42	(0.8895847797393799, 6.406432657968253e-05)	likely not a normal distribution
43	(0.9312201142311096, 0.002471240935847163)	likely not a normal distribution
44	(0.9208087921142578, 0.0009219571948051453)	likely not a normal distribution
45	(0.9086765646934509, 0.00031169259455055)	likely not a normal distribution
46	(0.9189855456352234, 0.0007799554732628167)	likely not a normal distribution
47	(0.9178503751754761, 0.0007033736910670996)	likely not a normal distribution
48	(0.9082732796669006, 0.00030099175637587905)	likely not a normal distribution
49	(0.8792779445648193, 2.8828462745877914e-05)	likely not a normal distribution
50	(0.9018280506134033, 0.00017381517682224512)	likely not a normal distribution
51	(0.9171023964881897, 0.000657280907034874)	likely not a normal distribution
52	(0.9022690057754517, 0.00018037116387858987)	likely not a normal distribution
53	(0.9297731518745422, 0.002147843362763524)	likely not a normal distribution
54	(0.9270318150520325, 0.0016514223534613848)	likely not a normal distribution
55	(0.9189062714576721, 0.0007743315654806793)	likely not a normal distribution
56	0.9336336255073547, 0.0031300410628318787)	likely not a normal distribution
57	(0.9233935475349426, 0.0011718327878043056)	likely not a normal distribution
58	(0.822623610496521, 6.148829356789065e-07)	likely not a normal distribution
59	(0.8976409435272217, 0.0001227615139214322)	likely not a normal distribution
60	(0.9102421998977661, 0.00035719736479222775)	likely not a normal distribution
61	(0.8996233940124512, 0.0001446087408112362)	likely not a normal distribution
62	(0.900181770324707, 0.00015147805970627815)	likely not a normal distribution
63	(0.8933776021003723, 8.675464050611481e-05)	likely not a normal distribution
64	(0.888426661491394, 5.8459794672671705e-05)	likely not a normal distribution
65	(0.9226863980293274, 0.0010970671428367496)	likely not a normal distribution
66	(0.9217154383659363, 0.0010025064693763852)	likely not a normal distribution
67	(0.9144634008407593, 0.0005185501649975777)	likely not a normal distribution
68	(0.8607951402664185, 7.490742518712068e-06)	likely not a normal distribution
69	(0.9234780669212341, 0.0011811222648248076)	likely not a normal distribution
70	(0.9273205995559692, 0.0016974825412034988)	likely not a normal distribution
71	(0.9182330369949341, 0.0007282581645995378)	likely not a normal distribution
72	(0.8516390323638916, 3.982294401794206e-06)	likely not a normal distribution
73	(0.8884052038192749, 5.836125637870282e-05)	likely not a normal distribution
74	(0.91771399974823, 0.0006947212386876345)	likely not a normal distribution

75	(0.8272948861122131, 8.202584353966813e-07)	likely not a normal distribution
76	(0.8161630034446716, 4.157644468705257e-07)	likely not a normal distribution
77	(0.8952314257621765, 0.00010080596985062584)	likely not a normal distribution
