A Personalized Visual Aid for Selections of Appearance Building Products with Long-term Effects

Chuhan Shi
cshiag@connect.ust.hk
The Hong Kong University of Science and Technology
Hong Kong, China

Zhihan Jiang
zhjiang@connect.hku.hk
The University of Hong Kong
Hong Kong, China

Xiaojuan Ma
mxj@cse.ust.hk
The Hong Kong University of Science and Technology
Hong Kong, China

Qiong Luo
luo@cse.ust.hk
The Hong Kong University of Science and Technology
Hong Kong, China

ABSTRACT

It is challenging for customers to select appearance building products (e.g., skincare products, weight loss programs) that suit them personally as such products usually demonstrate efficacy only after long-term usage. Although e-retailers generally provide product descriptions or other customers' reviews, users often find it hard to relate to their own situations. In this work, we proposed a pipeline to display envisioned users' appearance after long-term use of appearance building products to deliver their efficacy on each individual visually. We selected skincare as a case and developed SkincareMirror which predicts skincare effects on users' facial images by analyzing product function labels, efficacy ratings, and skin models' images. The results of a between-subjects study (N=48) show that (1) SkincareMirror outperforms the baseline shopping site in terms of perceived usability, usefulness, user satisfaction and helps users select products faster; (2) SkincareMirror is especially effective to males and users with limited product domain knowledge.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**.

KEYWORDS

Appearance building products, personalized visual aid, decision-making, virtual try-on

ACM Reference Format:
Chuhan Shi, Zhihan Jiang, Xiaojuan Ma, and Qiong Luo. 2022. A Personalized Visual Aid for Selections of Appearance Building Products with Long-term Effects. In *CHI Conference on Human Factors in Computing Systems (CHI '22), April 29-May 5, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3491102.3517659

1 INTRODUCTION

As human society places particular value on physical characteristics of an individual, there is a growing demand for products and services that can help people improve physical appearance and health [34], such as skincare products, dietary supplement medication, and weight loss programs. It is difficult for many consumers to select suitable products for physical appearance enhancement when they are exposed to a great variety of available candidates, especially when they lack knowledge in the related domains. Many popular e-commerce sites (e.g., Amazon[1], Taobao[2]) thus provide information such as product descriptions and brand stories as well as other customers' ratings and reviews [67] to facilitate customers' selection process. Nevertheless, reading such information is time-consuming and tedious for customers. Also, consumers always find it difficult to relate others' cases to their own situations, since long-term appearance building is highly personalized and most of them may not have enough product domain knowledge to distill useful information specific to themselves [14].

Similar issues in the fashion industry have driven the development of virtual try-on technologies, but these technologies mostly assist in selecting products that show effects immediately upon application, such as makeup [15] and clothes [28]. These virtual try-on services mainly consider the visual information that could be obtained from the appeal of the products, e.g., color and shape, and then directly render the known visual effects of a given product on top of user' face or body captured by a camera [23]. Such visual effects are typically stable and consistent across customers [39]. In contrast, the efficacy of long-term appearance building products is not visually observable upon application and usually varies from person to person after prolonged usage [61]. The final effects specific to each user, if visible, are determined by certain non-visual aspects of the products (e.g., ingredients) and individual customer's actual concerns (e.g., skin condition and body composition). Few works thus far have investigated the user experiences and perceptions of virtual try-on applications for products and services that take time to work.

This paper proposed a personalized visual aid to deliver the projected efficacy of the appearance building products on each individual visually by predicting the envisioned user appearance after long-term use of them. We first conducted a need-finding

[1]https://www.amazon.com
[2]https://www.taobao.com

CHI '22, April 29-May 5, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-9157-3/22/04…$15.00
https://doi.org/10.1145/3491102.3517659

study by interviewing 12 consumers of such products who have prior experiences with virtual try-on in other domains. We gained an understanding of their experiences, barriers, and needs when choosing long-term products suitable for themselves. We further summarized their specific requirements for the personalized visual aid to address their needs: locating the specific concern(s) of each individual, facilitating intuitive product comparison centered around the identified concern(s), and integrating all related information that helps make sense of envisioned product effects in consumers' decision-making process. Based on these design requirements, we designed a pipeline to mine efficacy-related product information (i.e., product descriptions/efficacy labels, customer reviews, images of models or end-users before and after using a product as shared on the official brand websites, etc.) and translated these information to parameters that control the visual effects of image processing algorithms that predict and render users' future physical appearance. As most participants in our need-finding study referred to skincare as an example when asked about their usage of appearance building products with long-term effects, we chose skincare as a case and developed SkincareMirror (Figure 1) based on our pipeline.

Rather than a full-fledged system for skincare product selection, SkincareMirror serves as a research prototype to explore whether such a personalized visual aid, when embedded in an e-shopping platform, can facilitate the inspection and interpretation of such products, and how it may influence customers' behavior and perception. To this end, we invited 48 participants to a between-subjects user study with the control group using the same experimental platform without our visual aid. Results from analysis of objective and subjective data captured in the experiment showed that with SkincareMirror, users explored more products in significantly shorter time than the control group did. The site with SkincareMirror significantly outperformed the baseline on perceived usability, usefulness, user satisfaction, and perceived informativeness. Motivated by the findings in prior literature that gender and experience differences exist in information processing during online shopping [24, 38], we further compared the behavior and perception of the participants with different genders and with different levels of skincare knowledge. Unlike the conclusion of previous work that male customers tend to use virtual makeup try-on as a playful app for enjoyment rather than as a tool supporting decision-making [30], we found that males and users with less skincare knowledge considered SkincareMirror to be an effective tool for their skincare product selection more than the female users and those more knowledgeable. Based on these findings, we discussed the challenges and concerns in designing personalized visual aids in appearance enhancement area, and design opportunities to further improve such personalized visual aids and enhance user experience.

In summary, the main contributions of this paper are as follows:

Figure 1: SkincareMirror shows a user's appearance before using a certain skincare product (left) and envisioned appearance after using the product (right).

(1) To benefit the selection process of appearance building products, we proposed a pipeline to develop visual aid to show envisioned users' appearance after using the appearance building products that aim for long-term effects.
(2) To showcase the proposed pipeline, we developed SkincareMirror based on it and conducted a user study to evaluate how the proposed visual aid, when embedded in a shopping website, impacts users' skincare selection perception and behavior. Both quantitative and qualitative analysis suggested that SkincareMirror can enhance consumer experience effectively compared with the baseline website without it. It is especially effective for males or users with limited skincare knowledge.
(3) We further proposed design opportunities to improve our pipeline and to address the difficulties in evaluating product effects that customers may face when selecting appearance building products with long-term effects.

2 RELATED WORK

2.1 Product Selection Supporting for Appearance Building Products with Long-term Effects

Recent work proposed various methods to provide product selection supporting for appearance building products with long-term effects. Many review-sharing websites, such as @cosme[3], sprang up. Yuki et al. presented a tag recommendation method for these sites helping filtering reviews and understanding the effects of products [44]. But the large amount of accessible information on such websites prevents customers from discovering the information they need [2]. Even if they can find correctly, they still have to spend a long time searching and studying all opinions [56].

Conventional recommendation systems are based on user profile (e.g., age) or/and item attributes (e.g., ingredients). For example, [36] recommended skincare through calculating the similarity of product ingredient composition. Nakajima et al. recommended skincare products that provide the desired effects based on users' age and skin type as well as the ingredients of products [50]. Other systems recommended products according to contextual information, such as [59] which provided long-term personalized support for healthy nutrition decisions by analyzing users' everyday dietary intake and physical activities. Also, some companies (e.g., Lemonbox[4]) began to produce tailored plans for various supplements

[3]https://www.cosme.net
[4]https://lemonbox.com.cn/

based on customers' input personal nutrition goals and lifestyle habits. As physical characteristics are also keys to determining the suitability of products, image-based systems have emerged. These systems require users to upload their personal photos, and then assist them in finding appropriate products based on users' characteristics captured from the photos. For instance, many applications like TroveSkin [70] offer skin analysis services and recommend skincare products based on the detected skin problems. Maki et al. utilized users' uploaded images to record their skin conditions and shared these data with professional skincare experts so that users can obtain appropriate skincare advice from the experts remotely [49]. In addition, many other product features customers may concern about were considered in previous works to support the selection of appearance building products with long-term effects, such as product price, side effects, availability of goods, and packaging [51].

Even though all existing works suggested possible appropriate products, they can not satisfy users' needs to try products and view the effects intuitively. Customers are prone to be frustrated when evaluating whether the final effects of the recommended products would satisfy their needs. In addition, it is often the desire to enhance appearance that drives customers to use appearance building products with long-term effects [41]. Predicting such products' possible outcomes and showing the changes of customers' appearances thus is important and needed.

2.2 Virtual Try-on

Virtual try-on technologies are widely used as they offer an opportunity to virtually try products before purchase [62]. Traditionally, based on computer graphics techniques, products' visual effects can be simulated by rendering the given product effects on top of users' faces or bodies in the output 2D images or 3D models [17, 21]. For example, Tong et al. used the quotient of the exemplar image "before" makeup divided by the exemplar image "after" makeup to indicate the makeup effects and transferred the effects to the target face to achieve virtual makeup try-on [69]. Guo et al. further completed the virtual makeup try-on with only the image "after" makeup as an exemplar by directly transferring the facial skin and color information in it to the subject image. With the burgeoning development of deep learning, researchers began to use deep learning models (e.g., GAN [20]) to synthesize the images of try-on products and target customers to predict the envisioned effects [22, 23, 37]. For instance, Yang et al. proposed an Adaptive Content Generating and Preserving Network which can achieve photo-realistic virtual clothes try-on by synthesizing a target clothing image and a reference human image while preserving details such as characteristics of clothes [81]. Liu et al. applied optimization-based neural makeup transfer models to synthesize the makeup of the reference face on the image showing the before-makeup face of the target user [40]. To provide more immersive shopping experience, virtual reality (VR) and augmented reality (AR) systems also emerged to a new trend. Eisert et al. proposed a real-time visualization in a virtual mirror environment by retexturing the garment dynamically [25]. Various virtual fitting rooms/mirrors were also developed (e.g., ModiFace[5]), enabling virtual try-on of makeup, haircut, nail salon, and teeth whitening.

However, these virtual try-on services were mainly for products that show effects immediately and they could not be directly applied to appearance building products with long-term effects. This is because models employed by these services aim to generate stable visual effects that are consistent across customers [39]. For instance, a makeup foundation could produce the same skin color and concealing effects regardless of who applies it with what kind of skin concern. In contrast, an acne removal skincare product would only show effects on users who indeed have acne issues; it would not work for those with redness but not acne [1]. Even for the customers with acne, the actual visual effects after long-term usage of the same product on them may be different. For instance, users with mild acne problems are likely to have lighter post-acne marks than those with severe acne problems. In short, existing virtual try-on services for fashion products that show instant effects mainly considered the visual information of products (e.g., color, shape) [22, 23, 63], and they focused primarily on how to align and render the standard product effects on users' faces or bodies. On the contrary, illustration of the efficacy of a long-term appearance building product requires comprehensive consideration of visual and non-visual characteristics of the product as well as customers' actual concerns. Few works have explored the design and user experiences of personalized visual aids for this type of products.

3 NEED-FINDING STUDY

To understand consumers' decision-making process in purchasing appearance building products with long-term effects, their needs, and the hurdles they encounter in product selection process, we conducted a qualitative need-finding study.

3.1 Participants and Procedure

We recruited 12 participants with diverse academic backgrounds by word-of-mouth and online advertisement via social media. Their ages range from 21 to 30 (M = 24.3, SD = 2.19), which falls in the main age group using online shopping [16]. All participants use appearance building products with long-term effects and have the needs to purchase them in daily life. They all have prior experiences with virtual try-on services in domains including makeup (5), shoes (3), clothes (2), and cosmetic contact lenses (2). We specified this inclusion criterion because we hope the participants can suggest more specific requirements for our visual aid based on their prior interaction with virtual try-on. We acknowledge that our participants lack diversity in their virtual try-on experience. However, it is reported that a considerable percentage of e-commerce consumers are likely to have been exposed to virtual try-on applications [35, 86] due to their growing popularity (used by more than 84% of major online retailers according to [85]). Thus, we believe that our need-finding results generally reflect the needs of ordinary customers of long-term appearance building products. As previous literature suggested that there exist considerable gender differences in appearance-related purchase [8], we recruited participants of different genders. Six participants were self-declared as male (M)

[5] https://modiface.com/

and six were female (F). To ensure the generality of our findings, we included consumers who identified themselves as knowledgeable (KN, five people) about such product selection as well as those who self-reported to be unknowledgeable (UN, seven people).

We conducted semi-structured interviews with these participants. After signing the consent form, they were first invited to recall and describe their latest experience of selecting appearance building products with long-term effects, including but not limited to what product information they focused on to help make decisions, how they collected such information, how long they spent on the decision-making process, and what step(s) took the most time. We also asked them about the difficulties they faced when choosing products suitable for them and their needs for facilitating their product selection. Then we asked the participants to recollect their most recent experience of using virtual try-on for purchasing such products with long-term effects. The interview questions mainly covered what they used such systems for, how they interacted with and perceived such a service, whether, and if so, how such systems benefited their decision-making process. Finally, we invited the participants to envision what services they would like to have in the selection process for long-term appearance enhancement products, and what expectations and concerns they had for such services.

3.2 Findings

All participants acknowledged that choosing appearance building products that require persistent usage and evaluating whether the products were suitable for them personally were difficult and time-consuming. We summarize the key insights from the need-finding study below.

Seven participants reported that one main obstacle of selecting appearance building products was that they found it hard to **obtain a sense of the possible outcome of a product that takes time to work without applying it for a long period of time**. Although current product selection supporting strategies (e.g., recommendation systems) could facilitate product selection to some extent, most (8/12) participants complained that they were still confused about what concerns they have and whether the products target their concerns. Three participants additionally stressed that, compared to existing virtual try-on application scenarios, the (visual) effects of such products are usually subtle, which makes it even harder for users to evaluate the products. P2 (M, UN) reported that he often failed to realize in time that he had made wrong product choices. Nearly half of the participants (5/12) reflected that they had made such mistakes before and worried that applying wrong products may cause adverse reactions and be harmful to their health. Therefore, most participants (8/12) felt that they need to be more careful and made more efforts to choose such products than other products that do the job immediately (e.g., makeup, shoes), trying to lower the risk of wasting money on wrong purchases. For example, they *"tend to buy products they have used before"* (P10, F, KN) but this may *"lead to missing more suitable products* (P11, F, UN).

In addition, five participants mentioned that retailers tend to put many function labels on each product to make the products easier to be searched. Product descriptions thus become very much alike. When facing such similar product descriptions, participants found it challenging to distinguish which product is more effective. Eight participants suggested that it would be helpful if our personalized visual aid could enable them to **intuitively distinguish the possible differences of products' effectiveness.** Interestingly, seven participants indicated higher tolerance of the disparity between the actual effect and the envisioned result displayed by virtual try-on for long-term appearance building products than for products that can show obvious effects upon application. This is because, as P1 (F, KN) put it, *"rather than getting absolutely accurate prediction results, being able to compare different [projected] product effects is more important [for products that take a long time to show outcomes]"*.

Furthermore, six participants expressed that shopping websites usually cannot **provide all the product information they need in one place** and that they tend to get distracted by other information on the sites that are not related to product efficacy, such as services, product packaging, and delivery. To make informed decisions, participants would turn to various other online platforms to collect information. For example, some participants mentioned that they would turn to review-sharing sites or mobile applications to check the *"reviews and ratings towards product efficacy from customers who have used the products"* (P1, F, KN). Due to this, customers' product selection process is especially time-consuming and they *"easily get lost in the searching process"* (P6, F, KN).

3.3 Design Requirements

Based on the feedback from the need-finding study, we derived a set of design requirements for a visual aid to facilitate selection of appearance building products with long-term effects:

D1. Locating the specific concern(s) of each individual. The individual differences and subtleties in appearance concerns make it hard for customers to choose long-term appearance enhancement products suitable for themselves. Therefore, our visual aid should help users quickly and explicitly localize physical area(s) specific to their concern(s) and present the predicted product effects, if any, in a targeted manner.

D2. Facilitating intuitive product comparison centered around the identified concern(s). From the need-finding study, we found that customers are usually confused about the increasingly alike product descriptions. Thus, it is important to facilitate users to intuitively differentiate the effectiveness of products, especially the products whose claimed functions are similar.

D3. Integrating all related information that helps make sense of envisioned product effects in consumers' decision-making process. According to our need-finding study, customers often have to navigate between platforms to collect and integrate their needed information. The visual aid should provide comprehensive product information that would benefit consumers' product selection process in an integrated fashion without context switching.

4 PIPELINE DESIGN AND IMPLEMENTATION

Based on the three design requirements, we developed a pipeline to automatically predict users' appearance after using a given product for a necessary period of time (Figure 2). As most of the need-finding participants mentioned skincare products as an example of such products, we chose skincare as a case domain and developed SkincareMirror based on our pipeline to explore whether the proposed

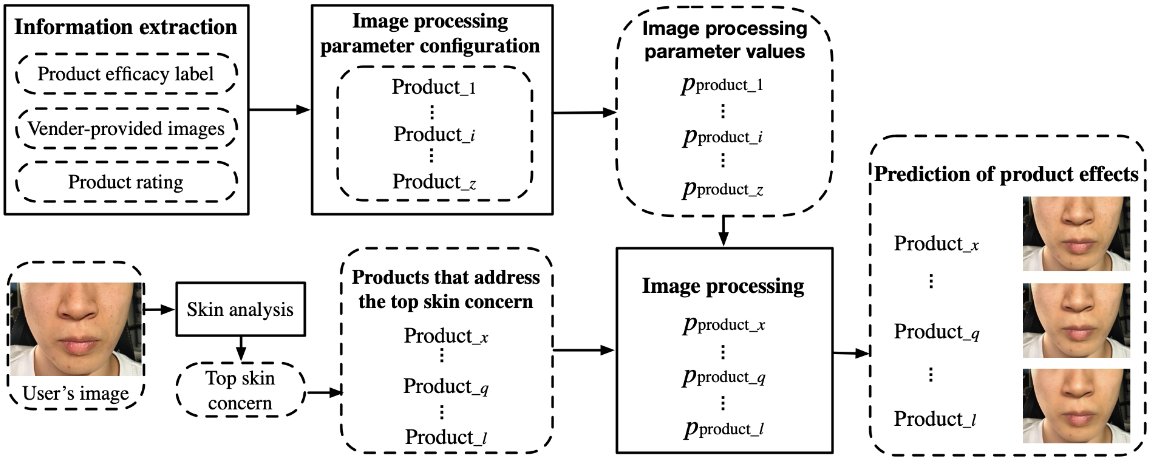

Figure 2: Pipeline flow. $\{\text{Product\_1}, \cdots, \text{Product\_}i, \cdots, \text{Product\_}z\}$ **and** $\{p_{\text{product\_1}}, \cdots, p_{\text{product\_}i}, \cdots, p_{\text{product\_}z}\}$ **denote the products in our dataset and their corresponding image processing parameter values.** $\{\text{Product\_}x, \cdots, \text{Product\_}q, \cdots, \text{Product\_}l\}$ **and** $\{p_{\text{product\_}x}, \cdots, p_{\text{product\_}q}, \cdots, p_{\text{product\_}l}\}$ **denote the products targeting user's top skin concern and their corresponding image processing parameter values.**

visual aid is beneficial for inspecting and interpreting products and how it may influence customers' product selection process.

4.1 Information Extraction

The information extraction module derives product efficacy labels, customer ratings, and pairs of images from real usage cases before (**Before-Image**) and after (**After-Image**) using products **(D3)**.

4.1.1 Collection of Product Efficacy Label and Consumer-provided Images. SkincareMirror provides prediction for four kinds of skincare products including (1) cream and moisturizer, (2) face treatment, (3) toners, and (4) eye treatment. We searched 3,703 products and their corresponding efficacy labels on Taobao.com, one of the most widely used shopping websites in China, and only 391 products were with Before- and After-Images on the official brand web pages. To describe product efficacy, Taobao offers 20 efficacy labels for each kind of skincare products, such as "Dark Circles" for eye treatment. Following these labels, the products of each kind can be divided into 20 efficacy categories. However, as most products each have multiple labels, each product may fit into multiple efficacy categories. Also, we noticed that some product labels have similar meanings. To remove redundancy, we randomly sampled 40 products under each label and merged two efficacy labels if the overlap of the items under these two labels is more than 80%. For instance, we randomly sampled 40 products of cream and moisturizer with the efficacy label "Hydration" and 40 with the label "Moisturization", respectively. We found the two samples have 34 products (i.e., 85% out of 40) in common. Therefore, we merged these two efficacy labels into a new label "Hydration and Moisturization". We kept the original labels if the products under the efficacy label overlap less than 80% with those under any other label. Finally, we obtained seven new efficacy labels for cream and moisturizer, eight for face treatment, six for toners, and four for eye treatment (details see Table 1). We then replaced products' original efficacy labels from Taobao with the corresponding new labels to describe the efficacy of each product in our dataset. To further simplify the selection process and avoid customer confusion, if a product had multiple new labels, we kept only one label that represented the primary efficacy of the product and put the product in the corresponding efficacy category.

4.1.2 Collection of Product Rating. After reassigning the efficacy label of each product, we collected product ratings as the measure of product effectiveness **(D2)**. Because Taobao does not limit users to rate only for efficacy, the ratings may be affected by many factors, such as packaging and services. Therefore, we collected ratings from "You are so beautiful today" [9], a popular mobile application that covers more than 400,000 products, has been used by more than 100 million people, and contains 5 million customer reviews. The ratings from this application are claimed to all come from users who have applied the products, and each product's rating will be shown only if there are more than ten reviewers on the product. In the app, users can rate a product from 0 to 10 (0 – no effect at all and 10 – perfect effect). We found that the ratings of products are all between 6.0 and 9.0, so we further divided the products in each efficacy category into three levels, i.e., [6.0, 7.0], (7.0, 8.0], and (8.0, 9.0]. For example, a facial cream with a primary efficacy label of "Hydration and Moisturization" and a user rating of 7.8 belongs to the "Hydration and Moisturization" efficacy category - (7.0, 8.0] level. As we created seven efficacy categories for cream and moisturizer products (Section 4.1.1) and each had three levels, we finally obtained 7×3 efficacy-level categories of this kind of products. Similarly, we got 8×3 efficacy-level categories of face treatment, 6×3 of toners, and 4×3 of eye treatment.

To have the same number of products with officially provided Before- and After-Images in each category to do the following image processing parameter configuration, we selected five such products in each efficacy-level category, which resulted in a total of 375 products with official images in our dataset.

Table 1: Image processing algorithms for different efficacy labels.

Skincare Product	Efficacy Label	Algorithms
Cream and Moisturizer	Hydration & Moisturization	Band-Sifting Decomposition (make skin look wetter) [4]
	Oil Control & Acne removal	Band-Sifting Decomposition (make skin look more drier) [4] & Bilateral Filter [19, 65, 68]
	Brightening	Band-sifting Decomposition (make skin look brighter and more glowing) [4]
	Lifting & Firming	Bilateral Filter [19, 65, 68]
	Pigmentation & Dark Spots Diminishing	Lookup Table Based Skin Tone Adjustment [57]
	Pore Reducing & Blackhead Removal	Bilateral Filter [19, 65, 68]
	After Sun Repair	Lookup Table Based Skin Tone Adjustment [57]
Face Treatment	Hydration & Moisturization	Band-Sifting Decomposition (make skin look wetter) [4]
	Blemish Removal	Bilateral Filter [19, 65, 68]
	Brightening	Band-sifting Decomposition (make skin look brighter and more glowing) [4]
	Acne Removal	Bilateral Filter [19, 65, 68]
	Redness Relieving	Lookup Table Based Skin Tone Adjustment [57]
	Lifting & Firming	Bilateral Filter [19, 65, 68]
	Pigmentation & Dark Spots Diminishing	Lookup Table Based Skin Tone Adjustment [57]
	Pore Reducing & Blackhead Removal	Bilateral Filter [19, 65, 68]
Toners	Hydration & Moisturization	Band-Sifting Decomposition (make skin look wetter) [4]
	Brightening	Band-sifting Decomposition (make skin look brighter and more glowing) [4]
	After Sun Repair	Lookup Table Based Skin Tone Adjustment [57]
	Pore Reducing & Blackhead Removal	Bilateral Filter [19, 65, 68]
	Even Skin Tone	Bilateral Filter [19, 65, 68]
	Oil Control & Acne removal	Band-Sifting Decomposition (make skin look more drier) [4] & Bilateral Filter [19, 65, 68]
Eye treatment	Hydration & Moisturization	Band-Sifting Decomposition (make skin look wetter) [4]
	Dark Circle Removal	Lookup Table Based Skin Tone Adjustment [57]
	Lifting & Firming & Eye Bag Removal	Bilateral Filter [19, 65, 68]
	Fat Granule Removal	Bilateral Filter [19, 65, 68]

4.2 Image Processing Parameter Configuration for Each Product

We applied different algorithms to produce the envisioned visual effects of skincare products of different efficacy. Details are shown in Table 1. For instance, we employed the bilateral filter algorithm to remove skin blemishes and smooth skin, because it is good at smoothing details while preserving edges [19] and is proven to have good performance in skin smoothing [10, 65, 82]. While we used band-sifting decomposition algorithm [4] to adjust the wetness and gloss of skin. Furthermore, as the skincare products are usually applied on different areas of a face, we constrained the effects of the image processing algorithms within the corresponding region(s) of the facial image. More specifically, for cream and moisturizer, face treatment, and toners which are commonly used on the whole face, we applied their associated algorithms on the entire facial skin area, while we would only apply the related algorithms on the eye areas for eye treatment for the skin around the eyes.

To ensure that the image processing algorithms could predict and generate visual effects that match the efficacy of individual product, we need to obtain the specific values of algorithm parameters for each product item (Figure 3). For a product i with officially provided pair of skin model's Before-Image (I_i) and After-Image (I_i^*), we first processed its Before-Image I_i with the image processing algorithm (A) corresponding to its efficacy category (Table 1). A uses a set of n parameters $P_A = \{P^{(1)}, P^{(2)}, \cdots, P^{(n)}\}$ to control its degree of image processing. Each parameter has a set of possible values, i.e., $p^{(j)}$ is a value of $P^{(j)}$ and $p^{(j)} \in \{p_1^{(j)}, p_2^{(j)}, \cdots, p_{N_j}^{(j)}\}$, where N_j is the number of possible values of parameter $P^{(j)}$ ($j = 1, 2, \cdots, n$). We can get all possible combinations ($m = N_1 \times N_2 \times \cdots \times N_n$) of the values for the parameters in P_A, and then configure the algorithm A with each value combination to process I_i, respectively. For example, the k^{th} ($k = 1, 2, \cdots, m$) parameter value combination is denoted as $\{p^{(1,k)}, p^{(2,k)}, \cdots, p^{(n,k)}\}$, where for each parameter $P^{(j)}$, its value in this combination $p^{(j,k)} \in \{p_1^{(j)}, p_2^{(j)}, \cdots, p_{N_j}^{(j)}\}$. The result image that is generated by this k^{th} combination is denoted as R_k. Consequently, we produced a series of m result images $\{R_1, R_2, \cdots, R_m\}$. Then we calculated the cosine similarity – an effective measure in facial image comparison [27, 52] – between each result image and the After-Image (I_i^*), and found the one with the maximum similarity. That is

$$R_t = \text{argmax}\left\{\text{similarity}\left(R_t, I_i^*\right) | R_t \in \{R_1, R_2, \cdots, R_m\}\right\}.$$

Here R_t is the result image produced by the t^{th} parameter value combination and has the highest cosine similarity to the original After-Image I_i^*. We thus set the final parameter values for product i to be $p_{A_i} = \{p_{A_i}^{(1)} = p^{(1,t)}, p_{A_i}^{(2)} = p^{(2,t)}, \cdots, p_{A_i}^{(n)} = p^{(n,t)}\}$, and regarded the effects generated by p_{A_i} as the predicted effect of product i.

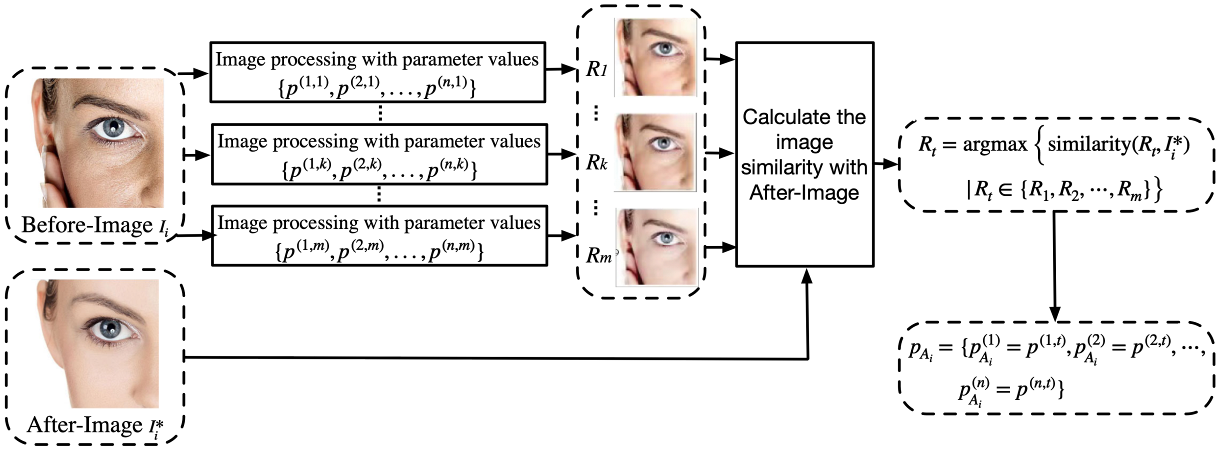

Figure 3: Flow of image processing parameter configuration for a product *i*.

In this way, for every efficacy-level category in our dataset, we obtained the parameter values for all products with Before- and After-Images in this category. Then, we take the median value of each parameter to configure the remaining products without skin model images. That is to say, assuming that product 1 to product l are all the items with Before- and After-Images in a given efficacy-level category, for any remaining product q in this category:

$$p_{A_q} = \Big\{\text{median}\{p_{A_1}^{(1)}, p_{A_2}^{(1)}, \cdots, p_{A_l}^{(1)}\}, \text{median}\{p_{A_1}^{(2)}, p_{A_2}^{(2)}, \cdots, p_{A_l}^{(2)}\}, \cdots, \text{median}\{p_{A_1}^{(n)}, p_{A_2}^{(n)}, \cdots, p_{A_l}^{(n)}\}\Big\}.$$

A pilot study was conducted to evaluate the results of image processing parameter configuration. We recruited seven participants with diverse academic backgrounds (4 female, 3 male; age range 22-26, M = 24.3, SD = 1.60) via word of mouth. They all have experience in using skincare products. We first collected their facial images (with consent) and the skincare products each of them used before. We identified the primary efficacy label and efficacy level of the most frequently used product for each participant, and then randomly chose two other products of the same kind and the same efficacy category but at two other efficacy levels. We processed the participant's facial image based on the corresponding image processing parameter values of the three products, and showed the three result images to the participant.

Next, we asked them to match images to efficacy levels to see whether they can easily and correctly distinguish the effects of products based on the images, because our need-finding results indicate that users may care more about whether the tool could allow them to compare product efficacy visually. Then we asked the participants whether the results were exaggerated, overprocessed, and unacceptable. Finally, we told them which image(s) corresponded to the products they had used before and asked whether the effects shown matched their experience. In general, all seven participants can easily distinguish the effects of the three levels of product efficacy. They all agreed that our image processing results were acceptable, matched their experience well, and would not cause any unrealistic expectations.

4.3 Implementation

Here we introduce how we implemented SkincareMirror to automatically predict product effects on each customer. After a user uploaded his/her facial image, we first applied the Meitu Skin Analysis API [45] to analyze skin concerns and detect area(s) of concerns in the images and determine the top skin concern by comparing each skin concern's area **(D1)**. Then we selected the products that can address the top skin concern from our product dataset based on efficacy labels. Next, we used the image processing parameter values of each product to process user-uploaded facial images and regarded the output as the prediction results on the product (Figure 2). As the effects of some skincare products are subtle, we highlighted the areas of the top skin concern using rectangles (Figure 4) to make it easier for users to see the changes on their facial images **(D1)**. For a product in our dataset that was not selected to address the user's top skin concern, we would simply return the user's original facial image to reduce the difficulties of distinguishing such products from those that can address user's other skin concerns but may only have subtle effects **(D2)**.

5 EXPERIMENT

In this section, we conducted a 2 (with SkincareMirror vs. without SkincareMirror) × 2 (male vs. female) × 2 (knowledgeable vs. unknowledgeable) between-subjects study to explore the answers to the following research questions:

RQ1: Whether (and how) would SkincareMirror affect users' skincare product selection behavior and perception?

RQ2: Whether (and how) would SkincareMirror affect users with different gender and knowledge in skincare?

RQ3: How would users perceive SkincareMirror, and what are their concerns?

To this end, we simulated an online shopping scenario and invited each participant to complete two separate tasks of choosing a skincare product (one about face treatment and the other about eye treatment) that is suitable for himself/herself. We counterbalanced the order of the tasks to alleviate potential order effects. In each task, participants were required to select only one product from 48 candidates. We randomly selected two products from each of 8×3 face treatment efficacy-level categories (Section 4.1.2) to get the

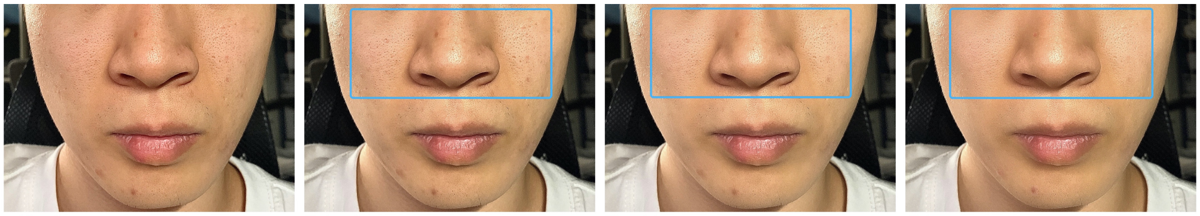

Figure 4: Example images with skin concerns on "Pores & Blackheads". (a) Original facial image. (b) level [6.0, 7.0]; image processing parameter=28. (c) level (7.0, 8.0]; image processing parameter=60. (d) level (8.0, 9.0]; image processing parameter=113. (*Note: The products may improve skin conditions on other facial area(s), we only highlight the area(s) of the top skin concern to clarify users' skin concerns and avoid misleading users*).

48 face treatment candidates and four products from each of 4×3 eye treatment efficacy-level categories (Section 4.1.2) to get the 48 eye treatment candidates. All relevant product data were collected from their brands' official websites. As for the participants' final decisions, there was no right or wrong answer [54]. We instead asked them to present the reasons for and basis of their decisions immediately after submitting their products of choice, to make sure that the participants take the tasks seriously. Here, we used the website without SkincareMirror rather than existing virtual try-on applications as the baseline because previous virtual try-on works are not applicable to the appearance building products with long-term effects.

5.1 Experimental Website Design

We developed two experimental websites with and without SkincareMirror respectively. To remove potential confounding effects in the between-subjects study, these two websites contained the same set of experimental stimuli. The website with SkincareMirror popped up an additional page for users to upload their facial photos with clear instructions and requirements of images. The main page of both experimental websites was a product catalog (Figure 5). These catalog pages, regardless of the condition, shared the same design modeled after Taobao and contained the same amount of information. More specifically, the catalog displayed 48 content tiles in a grid layout (Figure 5: (a)) and every tile (Figure 5: (c)) included an image, the name, the price, and a unique ID of a candidate product. Once participants have decided which product to choose, they can select the corresponding product ID (Figure 5: (c)) from the drop-down list (Figure 5: (b)) located at the bottom of the catalog page and submit their choices when ready. Clicking on a particular tile would take users to its associated item detail page (Figure 6). Item detail pages on both websites showed the basic product information (i.e., volume, expiration date, etc.) (Figure 6: (a)), detailed product description (i.e., main ingredients, usage method, etc.) (Figure 6: (c)), a pair of consumer-provided Before-Image and After-Image (Figure 6: (d)), as well as twenty customer reviews from brand official webpages on Taobao (Figure 6: (e)). The only difference between the two versions of experimental websites was that the one with SkincareMirror has an additional section presenting the user's original facial photo and the predicted user photo result generated by SkincareMirror on each item detail page between basic product information and detailed product description (Figure 6 : (b)). To minimize potential brand and price bias, we blurred all information about brands and kept all prices within a reasonable range.

5.2 Participants

We recruited 48 participants with diverse academic backgrounds (age range 18-31, M = 23.2, SD = 2.78) by word of mouth and online recruitment via social media. We randomly assigned half of the participants to the with-SkincareMirror group (SP1-24) and the other half to the without-SkincareMirror group (P1-24). All of our participants use skincare products in daily life and their average self-assessed knowledge in skincare product selection is 3.75 (SD = 1.60), with 1 being *no knowledge at all* and 7 being *a lot of knowledge*. To counterbalance users' gender and knowledge in skincare product selection, we recruited equal numbers of male (M) and female (F) participants and each gender had the same number of people who are knowledgeable (KN) and unknowledgeable (UN) (1~3 as unknowledgeable, 5~7 as knowledgeable, and we confirmed with participants who rated 4 again about their knowledge before assignment). As a result, in each group, there were 12 males (6 knowledgeable, 6 unknowledgeable) and 12 females (6 knowledgeable, 6 unknowledgeable). All participants are familiar with the Internet and have shopped online previously, meaning that they have no problem interacting with our experimental websites. Note that we only considered binary gender when exploring possible gender differences in this study following the practice of existing literature in related domain, so that we could compare our findings with prior works.

5.3 Measures Design

We designed eight measures (i.e., website usability, website usefulness, website credibility, user confidence in choice, website ability of improving user confidence in choice, user satisfaction with website, perceived informativeness, and trust of website) to measure users' skincare product selection perception (See Appendix A.1 for details). In addition, we asked participants in with-SkincareMirror group to rate their user trust of product effects showing on their own facial images (predicted by SkincareMirror), participants in without-SkincareMirror group to rate their trust of product effects showing on consumer-provided images, to investigate the difference. As for user behavior, we logged the user's total time spent on completing each task and the number of products a user clicked in each task. To further probe SkincareMirror influence user behavior in which part(s) of the experimental websites, we

Figure 5: Screenshots of product catalog page. (a) product catalog. (b) a drop-down list for users to submit their chosen products. (c) a tile of a candidate product.

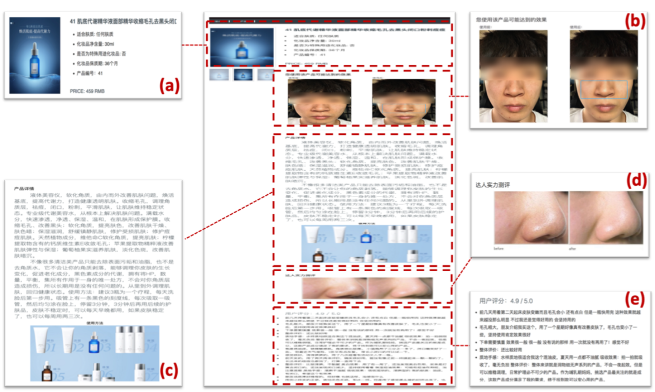

Figure 6: Screenshots of the item detail page with SkincareMirror. (a) basic product information. (b) SkincareMirror. (c) detailed product description. (d) consumer-provided Before-Image and After-Image. (e) customer reviews.

subsequently computed the user's time spent on scanning the basic product information, SkincareMirror's prediction images (only for with-SkincareMirror group), detailed product description, official skin model images, and customer reviews, respectively. (RQ1, RQ2)

To measure user perception towards SincareMirror, participants in with-SkincareMirror group also need to rate their user experience, adoption and use intention, in terms of SkincareMirror [55]. (RQ3)

5.4 Procedure and Data Collection

After obtaining participants' consent, we first asked them to complete a self-assessment on their knowledge in skincare product selection on a 7-point Likert scale. Then we introduced the procedure of the experiment and gave a quick tutorial on our visual aid to the with-SkincareMirror group. The tutorial was mainly about the functions of SkincareMirror and how to use it. We intentionally avoided any information which might cause bias such as names of service providers. For example, we said that we would detect users' skin concerns based on their facial images but do not reveal that it is based on Meitu Skin Analysis API [45]. In each task, participants can freely browse the given products. Once they have submitted their final choice, they need to write down the reasons for and basis of such a decision to complete the task. During this process, we ran a backend script to track and log each user behavior measure on the webpages. After finishing both tasks, participants were asked to rate their perception on a 7-point Likert scale ((1 – strongly disagree and 7 – strongly agree) regarding each perception measure respectively. To better understand participants' ratings and behavior, we further conducted a semi-structured interview with them. With users' consent, we recorded the audio of the whole interview. The whole experiment lasted for around 40 minutes. Upon its completion, the participants received a token of appreciation. The user study was approved by our institution's IRB.

5.5 Data Analysis

From the results of two-way ANOVA (between-subjects: with/without visual aid and task order, gender and task order, knowledge in skincare product selection and task order) tests, we confirmed that the task order has no significant main effect and interaction effect on every measure of participants' skincare product selection perception and behavior. Therefore, we combined the behavior data from the two tasks per person using the sum of data from the two tasks for each measure.

We planned to research not only the SkincareMirror's impact on participants but also the difference in SkincareMirror's impact on participants of different genders and the difference in SkincareMirror's impact on participants of different knowledge in skincare product selection. Therefore, we conducted two-way ANOVA on each behavior measure and perception measure regarding with/without SkincareMirror and gender, with/without SkincareMirror and knowledge, respectively as between-subjects variables. To probe the difference of participants with different genders and knowledge in perception and behavior towards SkincareMirror, we conducted one-way ANOVA analysis on user experience, adoption and use intention, and users' time spent on SkincareMirror with male/female and knowledgeable/unknowledgeable as between-subjects factors separately.

Finally, two authors conducted thematic analysis [5] on the transcripts of interview audio recordings and identified key themes in participants' feedback in response to our three research questions.

6 RESULTS

In this section, we present our findings about the possible impact of SkincareMirror on users' behavior and perception in skincare product selection, and further compare the effect between male and female users as well as between customers of different levels of knowledge about skincare product selection. We also discuss users' perception and concerns towards SkincareMirror.

6.1 Behavior in Skincare Product Selection Process (RQ1)

Comparing with the without-SkincareMirror group, SkincareMirror effectively shortens the total time of skincare product selection of the with-SkincareMirror group; ANOVA, $F(1,46) = 8.7$, $p < .01$,

Table 2: Means and standard errors of users' behavior in skincare product selection process.

	Without		With	
	M	**SE**	**M**	**SE**
Overall	882.33	583.31	487.43	302.90
Click	10.33	4.73	17.54	10.12
Details	390.60	337.95	124.29	101.94
Comments	328.20	215.71	131.69	127.04

Overall: the total time spent on the website (s). **Click**: the number of viewed products. **Details**: time spent on detailed product description (s). **Comments**: time spent on comments (s).

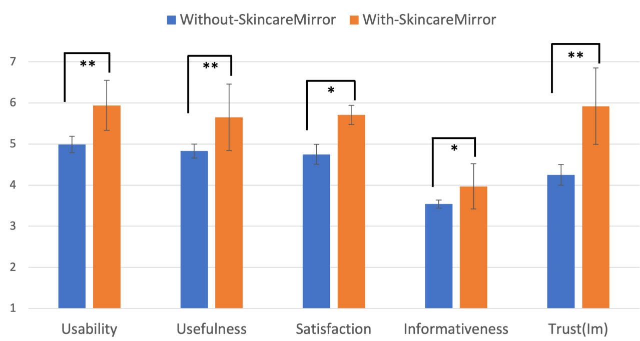

Figure 7: Means and standard errors of users' perception of website in terms of usability, usefulness, satisfaction, perceived informativeness, trust in images (∗: $p < .05$, ∗∗: $p < .01$).

$\eta^2 = .16$. In such shortened time, participants with SkincareMirror clicked and investigated significantly more products than those without SkincareMirror; $F(1,46) = 10.0$, $p < .01$, $\eta^2 = .18$ (Table 2). More specifically, the two sections that contain the most textual information are: detailed product description and customer reviews. Participants with SkincareMirror spent significantly less time on them (product description: $F(1,46) = 8.7$, $p < .01$, $\eta^2 = .23$; customer reviews: $F(1,46) = 14.8$, $p < .01$, $\eta^2 = .24$) than participants in without-SkincareMirror group (Table 2). These results show that, with SkincareMirror, users spend much less time on textual information. Consequently, our visual aid not only reduces the total decision-making time but also enables participants to view more products. On the time spent on basic product information and brand-provided Before- and After-Images of model faces, we find no significant difference between these two groups.

6.2 Perception in the skincare product selection Process (RQ1)

In general, SkincareMirror, when embedded in a shopping website, significantly improves the website's usability ($F(1,46) = 16.63$, $p < .01$, $\eta^2 = .27$), usefulness ($F(1,46) = 28.20$, $p < .01$, $\eta^2 = .38$), user satisfaction ($F(1,46) = 8.53$, $p < .05$, $\eta^2 = .16$), and perceived informativeness ($F(1,46) = 7.89$, $p < .05$, $\eta^2 = .15$) in product selection tasks (Figure 7). Although no significant effect is found on users' trust of the shopping website, participants' trust in SkincareMirror-provided preview images based on their own facial images is significantly higher than that in consumer-provided product effect images ($F(1,46) = 28.22$, $p < .01$, $\eta^2 = .38$) (Figure 7). In the rest of this section, we provide possible explanations of these results based on interview feedback from the participants.

6.2.1 Usability. Shopping websites for appearance building products with long-term effects usually tend to provide textual description of product efficacy as the visual effects of such products often vary for individuals, making it laborious to extract the key information (SP28, F, 26, KN) [3, 12]. We offer a visualization that directly shows users their facial images before and after virtually applying skincare products (SP16, M, 26, UN; SP24, F, 24, UN), providing them with an immediate understanding of the products' long-term effects. Nine participants explained that they could not fully understand the abstract product information or ingredient description in item detail pages. In comparison, SkincareMirror is easier for them to learn and to use. *"I hardly need to learn any extra skincare knowledge, just comparing some images is enough"* (SP3, M, 23, UN). This, however, may cause over-reliance on our visual aid, which we discuss in more details in the Discussion section.

6.2.2 Usefulness.

SkincareMirror makes up for users' lack of knowledge in skincare product selection. In advertisements or detailed product descriptions, it is common for customers to encounter unfamiliar terminologies that require abundant domain knowledge to understand [11]. This may mislead customers to have unrealistic expectations about the products' efficacy (P7, F, 23, KN), which may lead them to impulsive purchase without a comprehensive survey (P22, F, 20, UN). Another problem is that the facial skin condition of each individual is different [18] while users do not know their skin conditions well (SP23, F, 19, UN; SP3, M, 23, UN). So they do not know exactly how to evaluate whether the products are suitable for them. In our study, we find that participants feel it easier to make decisions with SkincareMirror, as it can *"automatically display visual effects and skin concerns' areas making inspecting products intuitively and easily"* (SP2, M, 23, UN) **(D1)**.

SkincareMirror greatly simplifies users' search for skincare information. Participants find it *"a struggling process to find useful and applicable information for evaluating the products"* (P14, F, 23, KN) although shopping websites show product information. 11 out of the 48 participants, including three users who are knowledgeable about skincare products, expressed that they are skeptical about product efficacy claimed by retailers (P5, F, 25, KN), other customers' reviews (P1, F, 28, UN), and sometimes recommendations from friends (SP17, F, 20, KN) for their skin conditions [13]. *"I think that a certain product works well for me, though I find others do not like it on review-sharing platforms"* (P4, M, 30, KN). With personalized prediction, SkincareMirror reduces users' burden on identifying information relevant to individual situation, by automatically performing skin analysis on users' facial images and envisioning possible product effects specific on them based on their skin concerns (SP14, F, 27, UN). This confirms that SkincareMirror meets our **D3**.

SkincareMirror helps participants distinguish the main functions of products. In order to increase sales, retailers tend to claim each product to have many functions [72]. Product descriptions thus become quite alike as they all put down more or less the same list of functions. In reality, a single product mainly solves one or two skin problems (P18, F, 19, KN). As a result, even if participants have different needs, there are many overlaps in the products, making them confused (P20, M, 24, UN). SkincareMirror helps with the product filtering, since we kept only one main verified function for each product in our database and only processed images of products that match the skin concerns in users' facial images **(D2)**. As the effects of skincare products are often subtle, the effective regions highlighted by colored frames make it easier for users to locate the area(s) of interests and pick out products whose main functions are suitable for their skin concerns (5/24) **(D1)**. *"Once I found the visual aid did not process my image [with no frames shown], I would directly turn to another product and would not look at the rest of the page any more"* (SP14, F, 27, UN).

6.2.3 Satisfaction. Results of the interview suggest that, participants show more satisfaction with the shopping website with SkincareMirror compared to the baseline. For example, SP18 (F, 18, UN) explicitly expressed her satisfaction that *"I really appreciate it [SkincareMirror]. Obviously, judging based on my own photos is better than based on others' cases"*, while no one in the group without the visual aid did so. The group with SkincareMirror also express more satisfaction with their final decisions than the group without in the interview. *"With the tool's help, I can easily tell that some products' effects are poor and some are good. I have an intuitive sense of the skincare products' effects, which makes me more sure of my choices"* (SP8, M, 24, UN) **(D2)**.

6.2.4 Perceived informativeness. Interestingly, although participants with SkincareMirror spent much less time on reading the textual information (Section 6.1), the survey results show that they perceived their decision to be more informative (of a significantly higher rating) than those in the baseline condition. On one hand, this is possibly because SkincareMirror reduces participants' burden of reading textual information and increases their *"interests to view more products"* (SP22, M, 23, KN). On the other hand, intuitive prediction of skincare effects makes participants dare to explore new products rather than conservatively choosing *"among popular products"* (P4, M, 30, KN) or *"from the products purchased before"* (SP7, M, 24, KN). SP17 (F, 20, KN) said that she found products she never knew before but thought more suitable for her than what she had been using. Especially, participants with limited knowledge about skincare products (7/24) explicitly commented that they could rarely get useful product information on shopping websites due to their lack of understanding of such information (e.g., unfamiliar terminologies). However, as they can easily understand and obtain insights from the visual results of SkincareMirror, they perceived that they gained more useful information. Nevertheless, some participants conveyed that showing ratings or processing parameters can help them understand the prediction results of SkincareMirror more accurately (SP2, M, 23, UN). In particular, participants who pay considerable attention to ingredients (SP4, F, 26, KN), especially those who are allergic to chemicals (SP3, M, 23, UN), stated that they cannot get what they want from SkincareMirror and they still need to search textual information from other sections on the website. For example, P7 (F, 23, KN) said *"I have to carefully check the ingredient list to make sure that there is not any ingredient that I am susceptible to, while SkincareMirror does not provide such hints"*. We discuss this in more details in the Discussion section.

6.2.5 Trust in images. An interesting finding from the interviews is that most participants (37/48) indicated that they do not trust others' photos of product effects on the Internet at all and seldom consider such photos during the selection process. One reason is that skin conditions vary from person to person and long-term skincare effects are highly personalized. Therefore, they doubted that *"retailers may only choose the best photos, but not everyone will achieve such efficacy"* (P5, F, 25, KN; SP10, F, 24, KN). In addition, *"many factors such as [color] brightness and image resolution may influence customers' perception"* (P13, F, 19, UN). In contrast, six of 24 participants in SkincareMirror condition said that they have more control over photos uploaded by themselves, so they believed the prediction result based on their own images more. Although our image processing system was based on models' photos of product effects, we only selected consumer-provided photos posted on official brand websites to control the quality, and avoided overprocessing images by using the median value of each efficacy-level category. Hence, SkincareMirror, which can visually display the differences between products for individual users, is deemed more trustworthy by our users. Four participants further stated that the trustworthiness of the organizations that provide SkincareMirror will also have an effect on them: they *"trust SkincareMirror more if it is developed by some well-known organizations that they are familiar with"*. In addition, as noted by SP11 (F, 23, KN) and SP17 (F, 20, KN), participants' trust is also established from the consistency of the prediction images with their previous actual experience using products with similar efficacy. However, seven participants expressed their concerns about the quality of our training data which may affect the effectiveness of our visual aid. *"Even if some customers may have applied the product, they might not have used it long enough before reviewing and rating"* (P5, F, 25, KN). Sometimes negative reviews and low ratings are not because the product is not good, but because *"reviewers have insufficient knowledge, which leads them to purchase unsuitable products"* (P12, F, 23, KN).

6.3 SkincareMirror's Effect on Users of Different Genders (RQ2)

The results of two-way ANOVA show that gender does not have a significant main effect on each behavior and perception measure. But it has significant interaction effects towards the total time spent on decision-making ($F(3,44) = 3.0$, $p < .05$, $\eta^2 = .17$), number of clicked products ($F(3,44) = 3.5$, $p < .05$, $\eta^2 = .19$), time spent on detailed product description ($F(3,44) = 4.4$, $p < .01$, $\eta^2 = .23$), and time spent on comments of the product ($F(3,44) = 5.6$, $p < .01$, $\eta^2 = .27$) (Table 3). The one-way ANOVA analysis shows that there is no significant effect in terms of the time spent on consumer-provided Before- and After-images, and on SkincareMirror (Table 3). As for perception, there are significant interaction effects towards credibility ($F(7,40) = 4.7$, $p < .05$, $\eta^2 = .23$) and satisfaction ($F(7,40) = 4.4$, $p < .05$, $\eta^2 = .10$) (Table 4). Although gender does not show a significant interaction effect in participants' confidence in their final choice (*"I am confident in my choice"*), there are significant interaction effects in participants' perceived SkincareMirror's ability

Table 3: Means and standard errors of behavior of male and female, knowledgeable and unknowledgeable group in with- and without-SkincareMirror group in terms of the total time spent on the website, the number of viewed products, time spent on basic product information, time spent on detailed product description, time spent on comments, time spent on consumer-provided Before- and After-Images, and time spent on SkincareMirror.

	Without				With				Without				With			
	Male		Female		Male		Female		Knowledgeable		Unknowledgeable		Knowledgeable		Unknowledgeable	
	M	SE	M	SE	M	SE	M	SE	M	SE	M	SE	M	SE	M	SE
Overall	834.4	144.8	930.3	194.7	425.0	86.3	549.9	88.5	789.8	111.8	984.9	211.9	501.9	100.3	473.0	76.8
Click	11.8	1.3	8.9	1.3	17.2	2.8	17.9	3.2	9.6	1.1	11.1	1.6	14.8	3.2	20.3	2.5
Basic	127.6	29.1	128.5	42.5	86.4	17.0	100.0	18.0	156.1	43.3	100.4	25.2	88.5	19.1	98.4	15.9
Details	386.0	78.8	395.2	117.0	101.6	30.5	147.0	28.0	284.5	47.0	496.7	125.1	133.0	33.5	115.9	26.0
Comments	282.5	46.1	373.9	74.9	107.2	44.3	156.2	27.1	316.5	61.5	339.9	65.6	166.2	43.9	97.2	25.8
Images	46.6	16.0	32.4	9.2	15.4	3.7	34.9	5.6	23.5	4.7	55.5	16.8	28.8	6.04	21.4	4.9
SkincareMirror	-	-	-	-	118.0	36.4	112.1	22.4	-	-	-	-	85.7	23.6	144.4	33.4

of improving their confidence (*"This website improves my confidence in my choices"*) (F(7,40) = 4.8, p < .01, η^2 = .04; Table 4). From the interview results, we find that one possible reason for the significant gender difference is that male participants prefer intuitive and interactive services offered by our tool which can help them decide quickly without spending much time on browsing the text, while females still refer to other information on the site and are more conservative. These are also consistent with prior findings about gender differences in information processing during online shopping, such as males are more affected by interactivity while females focus more on diagnostic information [38].

6.3.1 Behavior in Skincare Product Selection Process. In the interview, male participants made it clear that they prefer to see the images provided by SkincareMirror rather than read the complex product introduction. SP21 (M, 31, UN) reported that SkincareMirror would potentially change his habits of selecting skincare products, as SkincareMirror *"sparks the interest in viewing products because of its simplicity and enjoyment"*. In comparison, the main change in female participants' behaviors is that they would not collect information on other platforms as usual if they can use the website with SkincareMirror (SP17, F, 20, KN) and would not spend much time comparing similar products. *"I always hesitate among several products. Now I do not need to repeatedly compare these similar products, but make decisions directly based on the prediction"* (SP11, F, 23, KN). Also, SP19 (F, 21, KN) indicated that she would give priority to products that show better effects in SkincareMirror.

6.3.2 Credibility. In our study, most male participants (8/12) reported that they relied on the convenience brought by technologies in their daily life and their perceived credibility of SkincareMirror was mainly from such life experience. In comparison, females who think SkincareMirror is credible come to this assessment because SkincareMirror's simulated product effects match their actual experience with the products (5/12). Some participants (of both genders) shared the view that SkincareMirror's credibility would depend on the transparency of the data and technology used and/or the dependability of the service provider of SkincareMirror (SP6, M, 24, KN; SP5, F, 24, KN). During the interview, six participants asked for more details about the data and the image processing model. SP6 said that, *"I need to refer to official descriptions [of SkincareMirror], and further look into its data source and related algorithms."*. He further frankly reported that he *"would trust the visual aid only if the data and the models could be trusted"*. SP20 (M, 22, UN) implied that, *"If it is offered by some well-known companies or institutions, users would be more likely to believe in SkincareMirror"*.

6.3.3 Satisfaction. From the post-study interviews, we indeed find that the male participants show more satisfaction with and reliance on SkincareMirror, and their envision satisfaction after using the chosen products is higher than that of the females. *"I would be more satisfied with the products since I made decisions based on the personalized prediction of the products' effects rather than on my imagination built upon some online information that may not even be suitable for me"* (SP2, M, 23, UN). However, both male and female participants whose skin conditions are relatively good are not so satisfied with our visual aid. *"It is hard to tell the difference between after-images and original images since the changes are too subtle"* (SP24, F, 24, UN).

6.3.4 Ability of improving users' confidence. In the interview, male participants commented that the website with SkincareMirror could improve their confidence in their choices. They compared products based on the post-processing images (SP8, M, 24, UN; SP13, M, 24, UN), particularly when they are hesitant among several similar products (SP16, M, 26, UN). In contrast, five female participants said that many factors may influence products' final effects and they are still not very confident in their choices.

6.4 SkincareMirror's Effect on Users with Different Levels of Knowledge (RQ2)

Two-way ANOVA shows that knowledge in skincare product selection does not have a significant main effect on each behavior and perception measure. However, there are significant interaction effects towards total time spent on decision-making (F(3,44) = 3.2, p < .05, η^2 = .18), number of clicked products (F(3,44) = 4.4, p < .01, η^2 = .23), time spent on detailed product description (F(3,44) = 6.4, p < .01, η^2 = .30), and time spent on comments of the product (F(3,44) = 5.2, p < .01, η^2 = .26) (Table 3). But one-way ANOVA analysis shows that there is no significant effect in terms of the time spent on consumer-provided Before- and After-images, and on SkincareMirror (Table 3). Also, there are significant interaction effects towards users' perceived usefulness (F(7,40) = 3.4, p < .05, η^2 = .14) and confidence in their choices (F(7,40) = 3.9, p < .05, η^2 = .16) (Table 4).

6.4.1 Behavior in Skincare Product Selection Process. Our interview results show that knowledgeable and unknowledgeable participants

Table 4: Means and standard errors of perception of male and female, knowledgeable and unknowledgeable group in with- and without-SkincareMirror group in terms of website usability, website usefulness, website credibility, user confidence in choice,website ability of improving user confidence in choice, user satisfaction with website, perceived amount of information, trust of website, trust of images, user experience, and adoption and use intention.

	Without				With				Without				With			
	Male		Female		Male		Female		Knowledgeable		Unknowledgeable		Knowledgeable		Unknowledgeable	
	M	SE	M	SE	M	SE	M	SE	M	SE	M	SE	M	SE	M	SE
Usability	5.06	.21	4.92	.35	5.97	.17	5.92	.19	4.92	0.36	5.06	.18	6.08	.21	5.81	.13
Usefulness	4.44	.26	4.73	.31	6.06	.24	5.88	.23	4.43	.33	4.73	.23	5.90	.27	6.04	.19
Credibility	5.14	.19	5.53	.32	6.06	.25	5.31	.27	5.33	.29	5.33	.24	5.50	.30	5.86	.26
Confidence	5.08	.26	5.58	.34	5.83	.27	5.67	.14	5.67	.28	5.0	.30	5.75	.22	5.75	.22
Confidence (Ab)	4.33	.38	4.67	.33	6.33	.19	5.17	.42	4.50	.31	4.50	.40	5.83	.37	5.67	.38
Satisfaction	4.83	.32	5.08	.29	6.00	.25	5.67	.26	4.83	.32	5.67	.31	5.08	.29	6.00	.17
Informativeness	3.39	.14	3.69	.15	3.92	.15	4.03	.17	3.53	.17	3.56	.13	3.89	.17	4.06	.15
Trust	5.00	.21	5.42	.31	5.75	.28	5.17	.27	5.08	.34	5.33	.189	5.50	.29	5.42	.29
Trust (Im)	4.42	.29	4.08	.42	6.00	.25	5.83	.30	4.00	.21	4.50	.45	5.83	.30	6.00	.25

use SkincareMirror in different ways and that unknowledgeable participants have more behavior changes, as they rely more on SkincareMirror than the knowledgeable group. Specifically, four (of 12) unknowledgeable participants reported that they almost ignored the textual contents which they cannot understand and began to choose products solely based on SkincareMirror, whereas no one in the knowledgeable group reported this behavior. With SkincareMirror, unknowledgeable participants look at as many products as possible (SP3, M, 23, UN). When they find that there is no rectangle on the facial image indicating that the product is not suitable for them, they turn to the next product immediately. Therefore, the number of clicks they make is higher than the knowledgeable participants. The knowledgeable participants view only the products that they think may be suitable for them (SP12, M, 20, KN). Some knowledgeable participants, such as SP22 (M, 23, KN), also explicitly reported that SkincareMirror could prompt participants to click on more products to *"find lots of great products"* that they *"did not know before"*.

6.4.2 Usefulness. Unknowledgeable participants use SkincareMirror more during the selection process than knowledgeable participants, and some unknowledgeable participants *"almost completely rely on SkincareMirror to choose products"* (SP3, M, 23, UN). Especially, participants, who said that they never made skincare product selection decisions independently but absolutely followed others' suggestions, reported that they were willing and able to choose skincare products by themselves with the support of SkincareMirror (SP5, F, 24, UN).

6.4.3 Confidence. Participants in the unknowledgeable group tend to use SkincareMirror to select skincare products directly. They felt more confident in choosing skincare products based on SkincareMirror rather than based on other information on the website that they hardly understand (SP2, M, 23, UN). In contrast, participants in the knowledgeable group mostly said that they used SkincareMirror to do reconfirmation after they have made preliminary choices and that the visual aid thus further increased their confidence (SP7, M, 24, KN).

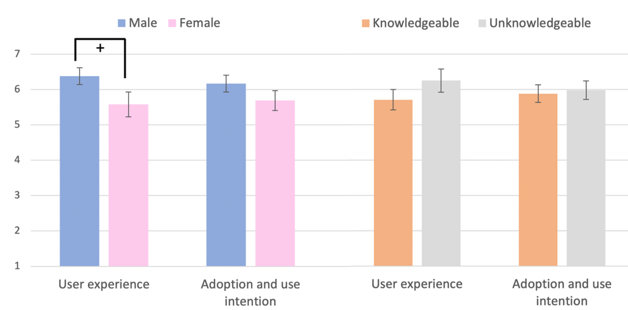

Figure 8: Means and standard errors of the perception of male and female, knowledgeable and unknowledgeable group towards SkincareMirror in terms of user experience, and adoption and use intention (+: $.05 < p < .1$).

6.5 Perception towards SkincareMirror and Concerns (RQ3)

6.5.1 User experience. Participants rate an average of 5.98 (SD = 1.09) on user experience about SkincareMirror, and gender has a marginally significant effect in one-way ANOVA analysis ($p = .07$) (Figure 8). Users praised the visual aid as it is *"attractive, interesting"* (SP6, M, 24, KN), *"intuitive"* (SP5, F, 24, UN), *"novelty, understandable"* (P1, F, 28, UN) and *"convenient"* (SP7, M, 24, KN). Participants also had positive comments on SkincareMirror's personalized results, *"it solves many problems in my daily skincare shopping process. I pay more attention to the photos rather than the other content on the page"* (SP8, M, 24, UN).

6.5.2 Adoption and use intention. In one-way ANOVA, neither gender nor knowledge in skincare selection shows a significant effect on visual aid adoption and use intention (Figure 8). Most participants intend to adopt SkincareMirror again, with an average rating of 5.93 (SD = 0.87). Participants reported that SkincareMirror improved their efficiency and was beneficial to decision-making. *"I am not afraid to choose skincare any more"* (SP2, M, 23, UN). 19 of 24 participants stated that they looked forward to continuing to use the visual aid in the future. SP22 (M, 23, KN) hoped SkincareMirror *"to be promoted to more platforms and be widely used"*. He added that *"it has broad prospects for development"*.

6.5.3 Privacy concerns. Surprisingly, only a few participants worry that SkincareMirror may leak their information with the uploaded photos of the entire faces. 25% participants are not concerned about the privacy issues at all, since *"uploading and scanning facial image is extremely common nowadays"* (SP20, M, 22, UN). More participants (33.3%) consider such concerns depending on the reputation and credibility of SkincareMirror and the organization providing the visual aid. Some participants suggested that it would be better if partial photos are allowed (SP21, M, 31, KN) and if participants can upload their face photos to local apps instead of online websites (SP14, F, 28, UN).

6.5.4 Concerns of photos without-makeup. Six of 24 participants indicated that they do not care about taking photos without makeup. Most participants regard uploading without-makeup photos as normal as *"without-makeup photos are necessary to detect skin concerns more accurately"* (SP14, F, 28, UN). *"My photos are to be processed by some algorithms, not to be publicly displayed."* (SP3, M, 23, UN). Another interesting finding is that those who are worried about this issue are all knowledgeable participants in skincare selection. Although the make-up trial is traditionally a female activity [30], we still find some male participants who worried about uploading photos without makeup or without beauty filters (SP6, M, 24, KN; P4, M, 30, KN).

7 DISCUSSION

This paper explores how users leverage a personalized visual aid that predicts their appearances after long-term application of appearance building products for decision-making. We selected skincare as a case and developed a research prototype called SkincareMirror to investigate users' experiences and concerns towards such a visual aid. While a prior study suggested that presenting to users individualized images of their future (aged) looks without using skincare products could cause discomfort [42], SkincareMirror is considered satisfactory, useful, and trustworthy (Section 6.2). It is possibly because the prior work intended to persuade customers to buy by showing a negative stimulus which may cause "loss aversion" [33, 87], while our goal is to support users' decision-making with an emphasis on the presence of a positive outcome, if any. Also, our approach makes users feel more in control by allowing them to upload personal facial images on their own (Section 6.2.5) instead of obtaining their information autonomously from certain sources [42, 43, 75] such as social media. This is in line with the previous finding in [6, 42] that knowing how their personal data is obtained has a comforting effect on users. Our visual aid is thus deemed more acceptable to consumers compared to the previous approaches (e.g., [42]) that also renders envisioned effects upon personal images.

From the user study, we find that SkincareMirror is perceived to be more effective in supporting decision-making for the male participants than the females. As discussed in Section 6.3, a possible reason is that when processing information during online shopping, male consumers benefit more from the interactivity of e-commerce platforms than females – the latter are more deliberate and prefer more diagnostic content [38]. However, our finding is not in line with the previous conclusion that male customers tend to use virtual try-on applications only for enjoyment rather than supporting their decision-making [30]. It may be because the usage scenario in the prior work is a cosmetics counter and the participants do not necessarily have the need for the makeup products [30]. In contrast, we only recruited participants who use skincare products in their daily life for our experiment. Further research is needed to investigate how our visual aid in public retail space might influence customer experience. It would be interesting to study how the tool appeals to people who never feel the need for appearance building and whether it could raise their interests and even purchase intents after "seeing" the potential long-term effects of products on them. In addition, we found that SkincareMirror is rated more effective by users with limited product domain knowledge than those who are knowledgeable. Interestingly, among the 12 participants who are familiar with skincare issues and products, four were reluctant to upload without-makeup/without-beauty-filter photos that show their actual skin conditions, whereas no unknowledgeable participants raised this concern (Section 6.5.4). It is probably because people who spend more time learning about makeup/skincare products are usually more concerned about their appearance; some might even have social appearance anxiety [47]. They, especially the latter group, may have less self-confidence showing non-idealized (e.g., without applying makeup and/or beauty filters) images of themselves [46]. Future work could explore ways to alleviate the concerns of such users, e.g., processing images and displaying the envisioned product effects locally rather than directly embedded on the websites [83], to make them feel more comfortable using personalized visual aids.

These findings uncover new aspects that the design of future personalized visual aids should take into account. Based on these insights, we discuss the generalizability of our work and propose the design considerations and opportunities for personalized visual aids of appearance building products that take time to work.

7.1 Generalizability of Our Work

Although we used skincare products as a case to illustrate our personalized visual aid design, the visual aid system could be easily extended to other appearance building products with long-term effects by simply using similar information of these products as the input data of our pipeline and applying the associated image editing algorithms to simulate the product effects. For example, to develop a personalized visual aid supporting the selection of weight loss programs, we can collect the data related to various weight loss programs, i.e., efficacy labels (e.g., arm fat loss, thigh fat loss, etc.), customer ratings, pairs of images from real usage cases before and after experiencing the programs, and feed them to our pipeline. Based on a user's weight loss goal(s), the pipeline could identify weight loss programs with efficacy label(s) that are related to the goal(s). Then, the corresponding part(s) of the user's body associated with these specific efficacy label(s) (e.g., arms, thighs, etc.) in the uploaded body images can be processed by the body shape deforming algorithms, such as Moving Least Squares [58], to predict the programs' effects on the user after persistent participation.

Besides, our findings that users with different genders and with different product knowledge used SkincareMirror in different ways have implications for the product selection support services for other appearance building products with long-term effects. We noticed that male participants praised the simplicity and interactivity of our tool, while females still wanted to refer to other product information on the website to complement SkincareMirror. Unknowledgeable participants relied on our tool to choose products, whereas those knowledgeable mostly used SkincareMirror after they have made preliminary decisions based on their experience. These findings stress the importance of tailoring the design of personalized visual aids (e.g., the product information integrated into visual aids) to users' genders as well as their prior knowledge and experiences in product selection.

7.2 Design Considerations and Opportunities

7.2.1 Analysis of physical characteristics and constitution. One of the design requirements (**D1**) derived from our need-finding study suggested that SkincareMirror should help users identify and locate their skin concerns and present the envisioned product effects associated with these specific concerns. However, diverse physical characteristics make the selection process of appearance building products with long-term effects particularly hard [18]. Users with special physical conditions and constitutions, especially those that are not visible on images such as allergic to certain chemical ingredients, may have special needs for product information (Section 6.2.4). To support such users, special visual effects could be displayed in our visual aid. For instance, we could ask users if they have a history of allergy to certain ingredients and, if so, add redness or rash on users' skin to alert them about the products that contain such ingredients. Furthermore, incorrect usage of appearance building products may lead to severe health consequences [8]. Because these effects may not show until after a long period, users may not realize them in time. In general, our pipeline could support selections of appearance building products with long-term effects further by identifying unsuitable products that users are searching for based on users' input images, e.g., someone who is too thin to use weight-loss products, and showing personalized projected visual effects of using the products to alert the user.

As we collected data from official brands and applied efficacy labels commonly used by skincare companies in our case, SkincareMirror's performance might be impacted by potential culture biases and social issues embedded in the practices of the beauty industry [79]. People of different cultural backgrounds have different physical characteristics and pursued appearance [32]. Hence, we could build customized visual aids by tuning efficacy labels and adding training images and ratings from people with diverse physical characteristics to adapt to different cultural contexts and different people.

7.2.2 Quality control of data. The quality of training data affects the effectiveness of our visual aid. In our user study, participants expressed their concerns about the quality of information (Section 6.2.5). Thus it is necessary to control the quality of product efficacy ratings and consumer-provided images to be used in our pipeline. For example, we could include the ratings only if they are given after a reasonable period from the time of purchase, or compute ratings based on customers' assessment of more specific and detailed product dimensions [29, 73]. We could also involve appearance building experts to help screen the images and ratings. In addition, we should try to reduce the potential bias regarding gender, age, culture, and other factors in the training data, ensuring that the collection well reflects the demographic distribution of the target population of such products [79].

7.2.3 Avoidance of over-reliance. We found that users wanted to process other types of product information on the website to complement SkincareMirror (Section 6.2.4). Their concern was that over-reliance on a single source of information may lead to biased decisions. To mitigate this issue, if we find that users are predominantly, and sometimes even exclusively, dependent on the visual aid, we could remind them of viewing other complementary information, provided that the users have agreed on receiving such reminders [54].

7.2.4 Ethical Issues. Many brands claim that their products have the function of whitening, which suggests that the whiter skin is more beautiful. Digital facial editing, as a widespread way of improving users' attractiveness, may affect people's setting of beauty standards in society. These influences make people's perceptions of beauty skew to exclude themselves. Their self-perception, self-esteem, and self-confidence may be negatively affected [84]. In recent years, inclusivity and diversity of beauty have come under the spotlight. Thus, we appeal to avoid using words such as whitening to describe product functions. Considering different human factors and offering customized digital facial editing effects for different people may make facial editing service friendly to more users and will encourage diverse beauty.

7.3 Limitations and Future Work

Our visual aid and experiments have several limitations. First, we use the same parameter value for image processing on the products that are in the same efficacy-level category but have no official model images. This treatment makes the effects of these products indistinguishable from each other to the users. Second, our visual aid does not report the length of time period to apply the corresponding product to have the envisioned effects. In reality, the visual effects of a product may develop over time [66]. As skincare products usually show effects after long-term usage, we have not examined users' post-purchase satisfaction for our study's time frame. Third, we model our experimental website after a well-established e-commerce platform and recruit a group of customers of the platform as participants. This sample may lack demographic diversity in terms of race and age, which may affect the generality of our experiment results. In addition, we only consider binary gender following existing literature in this domain. As a result, our gender-related findings may not adequately reflect the complexities of gender [7]. Furthermore, the actual (visual) effects of appearance enhancement products may be influenced by many factors, such as customers' current concerns, physique, lifestyle, and product efficacy. In the scope of this paper, we only consider the matching between users' skin concerns and product efficacy and do not collect the participants' physique and lifestyle information due to privacy issues.

In the future, we plan to take more kinds of appearance building products with long-term effects and physical characteristics into consideration and differentiate within the same efficacy-level category the effects of products that have no consumer-provided images so that users can easily recognize the differences of these products, if present, from the prediction results. In addition, we will further study users' post-purchase perception and explore users' experience of using our visual aid to select products for people other than themselves. We will also explore gender-related issues under a broader definition of gender. Moreover, we are interested in exploring the in-store application of such a service and seeing how it affects customers' offline shopping process and experience. In our user study, we do not use prior virtual try-on experience as an inclusion/exclusion criteria. It is possible that customers who never used virtual try-on prior to the experiment may experience "novelty effects" when using our personalized visual aid [26], while those who had been exposed to other virtual try-on services may encounter the Baby Duck Syndrome [60]. Therefore, the possible impacts of users' experiences of virtual try-on could also be a potential direction for future study.

8 CONCLUSION

Selecting suitable appearance building products that take time to show effects is especially challenging, since their efficacy is not visually observable upon application. Also, the product effects usually vary among individuals based on the individual's actual concerns and certain non-visual aspects of the products (e.g., ingredients). In this paper, we presented a pipeline to develop a visual aid that displays the envisioned users' appearance after using the appearance building products that target long-term effects. We chose skincare as a case and developed SkincareMirror which predicts skincare products' possible outcomes when applied to ones' skin based on their facial images. A between-subjects study suggested that SkincareMirror facilitates participants to survey more products in significantly less time. This demonstrates SkincareMirror's efficacy in shortening the decision-making process and encouraging users to inspect suitable products that are previously unknown to them. Also, participants perceived that SkincareMirror is useful for evaluating products and is easy to use. Their perceived informativeness from the shopping site and the degree of satisfaction are also boosted by our visual aid. Furthermore, we found that SkincareMirror is more effective for male and unknowledgeable participants in supporting skincare products selection than female and knowledgeable participants, respectively. Our findings have implications for future product selection support services in the physical appearance enhancement industry. This work is also an important step of extending personalized image work from the products whose efficacy is stable, consistent, and immediately noticeable to products whose efficacy is uncertain upon application. Based on the insights from this work, future studies could further explore the application of such a service in offline environments and its effects on a broader range (e.g., age) of user groups.

ACKNOWLEDGMENTS

Many thanks to the anonymous reviewers for their insightful suggestions. We thank Xingbo Wang and Yicheng Hu for their valuable inputs. This research was supported in part by HKUST-WeBank Joint Laboratory Project Grant No.: WEB19EG01-d.

REFERENCES

[1] I Bartenjev, L Oremović, M Rogl Butina, I Sjerobabski Masnec, A Bouloc, K Voda, and M Šitum. 2011. Topical effectiveness of a cosmetic skincare treatment for acne-prone skin: a clinical study. *Acta Dermatoven APA* 20, 2 (2011).

[2] Shlomo Berkovsky and Jill Freyne. 2015. Web personalization and recommender systems. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2307–2308.

[3] Carlos Flavián Blanco, Raquel Gurrea Sarasa, and Carlos Orús Sanclemente. 2010. Effects of visual and textual information in online product presentations: looking for the best combination in website design. *European Journal of information systems* 19, 6 (2010), 668–686.

[4] Ivaylo Boyadzhiev, Kavita Bala, Sylvain Paris, and Edward Adelson. 2015. Band-sifting decomposition for image-based material editing. *ACM Transactions on Graphics (TOG)* 34, 5 (2015), 1–16.

[5] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[6] Evert Van den Broeck, Karolien Poels, and Michel Walrave. 2017. A factorial survey study on the influence of advertising place and the use of personal data on user acceptance of Facebook ads. *American behavioral scientist* 61, 7 (2017), 653–671.

[7] Joy Buolamwini and Timnit Gebru. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*. PMLR, 77–91.

[8] Scot Burton, Richard G Netemeyer, and Donald R Lichtenstein. 1995. Gender differences for appearance-related attitudes and behaviors: Implications for consumer welfare. *Journal of Public Policy & Marketing* 14, 1 (1995), 60–75.

[9] C2H4. 2017. *You are so beautiful today.* Retrieved August 17, 2020 from http://c2h4.org

[10] Chih-Wei Chen, Da-Yuan Huang, and Chiou-Shann Fuh. 2009. Automatic skin color beautification. In *International Conference on Arts and Technology*. Springer, 157–164.

[11] Jen-Yi Chen. 2013. Advertisements of skin care products in Taiwan. *Mass Communicator: International Journal of Communication Studies* 7, 4 (2013), 19–25.

[12] Yu-Chen Chen, Rong-An Shang, and Chen-Yu Kao. 2009. The effects of information overload on consumers' subjective state towards buying decision in the internet shopping environment. *Electronic Commerce Research and Applications* 8, 1 (2009), 48–58.

[13] Yi-Fen Chen and Hui-Fen Lu. 2015. We-commerce: Exploring factors influencing online group-buying intention in T aiwan from a conformity perspective. *Asian Journal of Social Psychology* 18, 1 (2015), 62–75.

[14] Elizabeth Cowley and Andrew A Mitchell. 2003. The moderating effect of product knowledge on the learning and organization of product information. *Journal of Consumer Research* 30, 3 (2003), 443–454.

[15] Filipe Morgado Simoes de Campos and Carlos H Morimoto. 2014. Virtual makeup: Foundation, eye shadow and lipstick simulation. In *2014 XVI Symposium on Virtual and Augmented Reality*. IEEE, 181–189.

[16] DEM Dharshan and Japa Asritha Reddy. 2019. A Study of Customer Behaviour towards Online Shopping in Hyderabad. *International Journal of trend in scientific research and development* 3 (2019), 741–744.

[17] Jun Ehara and Hideo Saito. 2006. Texture overlay for virtual clothing based on PCA of silhouettes. In *2006 IEEE/ACM International Symposium on Mixed and Augmented Reality*. Citeseer, 139–142.

[18] Yuri Fitrian, Miftachul Huda, Adibah Muhtar, Azhar Yaakub Arifin, Norhafizah Musa, Kamarul Shukri Mat Teh, Novita Mulya Sari, and Andino Maseleno. 2019. Application design for determining suitable cosmetics with the facial skin type using fuzzy logic approach. *Journal of Computational and Theoretical Nanoscience* 16, 5-6 (2019), 2153–2158.

[19] Sanjay Ghosh, Ruturaj G Gavaskar, Debasisha Panda, and Kunal N Chaudhury. 2019. Fast Scale-Adaptive Bilateral Texture Smoothing. *IEEE Transactions on Circuits and Systems for Video Technology* (2019).

[20] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial networks. *arXiv preprint arXiv:1406.2661* (2014).

[21] Peng Guan, Loretta Reiss, David A Hirshberg, Alexander Weiss, and Michael J Black. 2012. Drape: Dressing any person. *ACM Transactions on Graphics (TOG)* 31, 4 (2012), 1–10.

[22] Xintong Han, Xiaojun Hu, Weilin Huang, and Matthew R Scott. 2019. Clothflow: A flow-based model for clothed person generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 10471–10480.

[23] Xintong Han, Zuxuan Wu, Zhe Wu, Ruichi Yu, and Larry S Davis. 2018. Viton: An image-based virtual try-on network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 7543–7552.

[24] Lee Hao Suan Samuel, MS Balaji, and Khong Kok Wei. 2015. An investigation of online shopping experience on trust and behavioral intentions. *Journal of Internet Commerce* 14, 2 (2015), 233–254.

[25] Anna Hilsmann and Peter Eisert. 2009. Tracking and retexturing cloth for real-time virtual clothing applications. In *International Conference on Computer Vision/Computer Graphics Collaboration Techniques and Applications*. Springer, 94–105.

[26] Wen Huang. 2020. *Investigating the novelty effect in virtual reality on stem learning*. Ph.D. Dissertation. Arizona State University.

[27] Sibt Ul Hussain, Thibault Napoléon, and Fréderic Jurie. 2012. Face recognition using local quantized patterns.

[28] Surgan Jandial, Ayush Chopra, Kumar Ayush, Mayur Hemani, Balaji Krishnamurthy, and Abhijeet Halwai. 2020. Sievenet: A unified framework for robust image-based virtual try-on. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2182–2190.

[29] Sirrka L Jarvenpaa and Peter A Todd. 1996. Consumer reactions to electronic shopping on the World Wide Web. *International Journal of electronic commerce* 1, 2 (1996), 59–88.

[30] Ana Javornik, Yvonne Rogers, Ana Maria Moutinho, and Russell Freeman. 2016. Revealing the shopper experience of using a" magic mirror" augmented reality make-up application. In *Conference on designing interactive systems*, Vol. 2016. Association for Computing Machinery (ACM), 871–882.

[31] Thomas J Johnson and Barbara K Kaye. 2004. Wag the blog: How reliance on traditional media and the Internet influence credibility perceptions of weblogs among blog users. *Journalism & mass communication quarterly* 81, 3 (2004), 622–642.

[32] Jaehee Jung and Seung-Hee Lee. 2006. Cross-cultural comparisons of appearance self-schema, body image, self-esteem, and dieting behavior between Korean and US women. *Family and Consumer Sciences Research Journal* 34, 4 (2006), 350–365.

[33] Daniel Kahneman and Amos Tversky. 2013. Prospect theory: An analysis of decision under risk. In *Handbook of the fundamentals of financial decision making: Part I*. World Scientific, 99–127.

[34] Mai Ngoc Khuong and Hoang Thi My Duyen. 2016. Personal factors affecting consumer purchase decision towards men skin care products—A study in Ho Chi Minh City, Vietnam. *International Journal of Trade, Economics and Finance* 7, 2 (2016), 44–50.

[35] Jiyeon Kim and Sandra Forsythe. 2008. Adoption of virtual try-on technology for online apparel shopping. *Journal of Interactive Marketing* 22, 2 (2008), 45–59.

[36] Gyeongeun Lee. [n.d.]. A Content-based Skincare Product Recommendation System. ([n. d.]).

[37] Tingting Li, Ruihe Qian, Chao Dong, Si Liu, Qiong Yan, Wenwu Zhu, and Liang Lin. 2018. Beautygan: Instance-level facial makeup transfer with deep generative adversarial network. In *Proceedings of the 26th ACM international conference on Multimedia*. 645–653.

[38] Xiaolin Lin, Mauricio Featherman, Stoney L Brooks, and Nick Hajli. 2019. Exploring gender differences in online consumer purchase decision making: An online product presentation perspective. *Information Systems Frontiers* 21, 5 (2019), 1187–1201.

[39] Guoqiang Liu, Dan Song, Ruofeng Tong, and Min Tang. 2021. Toward Realistic Virtual Try-on Through Landmark-Guided Shape Matching. (2021).

[40] Si Liu, Xinyu Ou, Ruihe Qian, Wei Wang, and Xiaochun Cao. 2016. Makeup like a Superstar: Deep Localized Makeup Transfer Network. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence* (New York, New York, USA) *(IJCAI'16)*. AAAI Press, 2568–2575.

[41] Elizabeth T Makino, Lily I Jiang, Priscilla Tan, Tsing Cheng, and Rahul C Mehta. 2018. Addressing male facial skin concerns: clinical efficacy of a topical skincare treatment product for men. *Journal of drugs in dermatology: JDD* 17, 3 (2018), 301–306.

[42] Miguel Malheiros, Charlene Jennett, Snehalee Patel, Sacha Brostoff, and Martina Angela Sasse. 2012. Too close for comfort: A study of the effectiveness and acceptability of rich-media personalized advertising. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 579–588.

[43] Aleksandar Matic, Martin Pielot, and Nuria Oliver. 2017. " OMG! How did it know that?" Reactions to Highly-Personalized Ads. In *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization*. 41–46.

[44] Yuki Matsunami, Mayumi Ueda, and Shinsuke Nakajima. 2017. Tag recommendation method for a cosmetics review recommender system. In *Proceedings of the 19th International Conference on Information Integration and Web-based Applications & Services*. 166–170.

[45] Meitu. 2015. *Meitu AI Open Platform Skin Analysis*. Retrieved August 23, 2020 from https://ai.meitu.com

[46] Fiona Monro and Gail Huon. 2005. Media-portrayed idealized images, body shame, and appearance anxiety. *International Journal of Eating Disorders* 38, 1 (2005), 85–90.

[47] Amy Muise and Serge Desmarais. 2010. Women's perceptions and use of "anti-aging" products. *Sex Roles* 63, 1-2 (2010), 126–137.

[48] Cosmin Munteanu, Ronald Baecker, Gerald Penn, Elaine Toms, and David James. 2006. The effect of speech recognition accuracy rates on the usefulness and usability of webcast archives. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. 493–502.

[49] Maki Nakagawa, Koji Tsukada, and Itiro Siio. 2010. Remote skincare advice system using life logs. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems*. 3811–3816.

[50] Yoko Nakajima, Hirotoshi Honma, Haruka Aoshima, Tomoyoshi Akiba, and Shigeru Masuyama. 2019. Recommender System Based on User Evaluations and Cosmetic Ingredients. In *2019 4th International Conference on Information Technology (InCIT)*. IEEE, 22–27.

[51] Nurliana Nasution, Bayu Febriadi, Galih Mahalisa, Nadiya Hijriana, Muhammad Rasyidan, Dewinta Marthadinata Sinaga, Sinta Maulina Dewi, Agus Perdana Windarto, Novita Aswan, and Mokhamad Ramdhani Raharjo. 2020. Application of ELECTRE Algorithm in Skincare Product Selection. *JPhCS* 1471, 1 (2020), 012066.

[52] Hieu V Nguyen and Li Bai. 2010. Cosine similarity metric learning for face verification. In *Asian conference on computer vision*. Springer, 709–720.

[53] Zhenhui Peng, Qingyu Guo, Ka Wing Tsang, and Xiaojuan Ma. 2020. Exploring the effects of technological writing assistance for support providers in online mental health community. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–15.

[54] Zhenhui Peng, Yunhwan Kwon, Jiaan Lu, Ziming Wu, and Xiaojuan Ma. 2019. Design and evaluation of service robot's proactivity in decision-making support process. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.

[55] Pearl Pu, Li Chen, and Rong Hu. 2011. A user-centric evaluation framework for recommender systems. In *Proceedings of the fifth ACM conference on Recommender systems*. 157–164.

[56] Pakawan Pugsee, Patcharaporn Sombatsri, and Rujiporn Juntiwakul. 2017. Satisfactory analysis for cosmetic product review comments. In *Proceedings of the 2017 International Conference on Data Mining, Communications and Information Technology*. 1–6.

[57] Frode Eika Sandnes, Levent Neyse, and Yo-Ping Huang. 2016. Simple and practical skin detection with static RGB-color lookup tables: A visualization-based study. In *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 002370–002375.

[58] Scott Schaefer, Travis McPhail, and Joe Warren. 2006. Image deformation using moving least squares. In *ACM SIGGRAPH 2006 Papers*. 533–540.

[59] Hanna Schäfer. 2016. Personalized support for healthy nutrition decisions. In *Proceedings of the 10th ACM Conference on Recommender Systems*. 455–458.

[60] P Seebach. 2005. Baby duck syndrome: Imprinting on your first system makes change a very hard thing. *IBM DeveloperWorks* (2005).

[61] Masami SENOO, Yuko TAKEMOTO, and Hideo JINGU. 2002. Change in affections by continuous use of skincare cosmetics. *Kansei Engineering International* 3, 3 (2002), 31–36.

[62] Anne R Smink, Sanne Frowijn, Eva A van Reijmersdal, Guda van Noort, and Peter C Neijens. 2019. Try online before you buy: How does shopping with augmented reality affect brand responses and personal data disclosure. *Electronic Commerce Research and Applications* 35 (2019), 100854.

[63] Dan Song, Tianbao Li, Zhendong Mao, and An-An Liu. 2020. SP-VITON: shape-preserving image-based virtual try-on network. *Multimedia Tools and Applications* 79, 45 (2020), 33757–33769.

[64] Brendan Spillane, Isla Hoe, Mike Brady, Vincent Wade, and Séamus Lawless. 2020. Tabloidization versus credibility: Short term gain for long term pain. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–15.

[65] Sanun Srisuk. 2014. Bilateral filtering as a tool for image smoothing with edge preserving properties. In *2014 International Electrical Engineering Congress (iEECON)*. IEEE, 1–4.

[66] STYLECASTER. 2018. *How Long Does It Really Take for Skin-Care Products to Deliver Results?* Retrieved Sep 2, 2020 from https://stylecaster.com/beauty/how-long-til-skin-care-products-work/

[67] Songsri Tangsripairoj, Kwanchanok Khongson, Pitchapa Puangnak, and Yada Boonserm. 2018. SkinProf: An Android Application for Smart Cosmetic and Skincare Users. In *2018 15th International Joint Conference on Computer Science and Software Engineering (JCSSE)*. IEEE, 1–6.

[68] Carlo Tomasi and Roberto Manduchi. 1998. Bilateral filtering for gray and color images. In *Sixth international conference on computer vision (IEEE Cat. No. 98CH36271)*. IEEE, 839–846.

[69] Wai-Shun Tong, Chi-Keung Tang, Michael S Brown, and Ying-Qing Xu. 2007. Example-based cosmetic transfer. In *15th Pacific Conference on Computer Graphics and Applications (PG'07)*. IEEE, 211–218.

[70] TroveSkin. 2017. *TroveSkin: A FREE Skincare App for Skin Analysis & Track*. Retrieved August 23, 2020 from https://www.troveskin.com

[71] April Yi Wang, Dakuo Wang, Jaimie Drozdal, Michael Muller, Soya Park, Justin D Weisz, Xuye Liu, Lingfei Wu, and Casey Dugan. 2021. Themisto: Towards Automated Documentation Generation in Computational Notebooks. *arXiv preprint arXiv:2102.12592* (2021).

[72] Fuxue Wang, Yi Li, and Yiwen Zhang. 2011. An empirical study on the search engine optimization technique and its outcomes. In *2011 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC)*. IEEE, 2767–2770.

[73] Lin Wang, Huijie Tang, Dong Liu, and Caiyan Xing. 2017. Study on Customer Experience Dimension Based on Online Shopping reviews of MI's Products. In *MATEC Web of Conferences*, Vol. 100. EDP Sciences, 05024.

[74] Xingbo Wang, Haipeng Zeng, Yong Wang, Aoyu Wu, Zhida Sun, Xiaojuan Ma, and Huamin Qu. 2020. VoiceCoach: Interactive Evidence-based Training for Voice Modulation Skills in Public Speaking. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12.

[75] Jeffrey Warshaw, Tara Matthews, Steve Whittaker, Chris Kau, Mateo Bengualid, and Barton A Smith. 2015. Can an Algorithm Know the" Real You"? Understanding People's Reactions to Hyper-personal Analytics Systems. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 797–806.

[76] Stephan Weibelzahl, Alexandros Paramythis, and Judith Masthoff. 2020. Evaluation of adaptive systems. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and personalization*. 394–395.

[77] Shaomei Wu, Lindsay Reynolds, Xian Li, and Francisco Guzmán. 2019. Design and evaluation of a social media writing support tool for people with dyslexia. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–14.

[78] Meng Xia, Mingfei Sun, Huan Wei, Qing Chen, Yong Wang, Lei Shi, Huamin Qu, and Xiaojuan Ma. 2019. Peerlens: Peer-inspired interactive learning path planning in online question pool. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–12.

[79] Qinwei Xie and Meng Zhang. 2013. White or tan? A cross-cultural analysis of skin beauty advertisements between China and the United States. *Asian Journal of Communication* 23, 5 (2013), 538–554.

[80] Litao Yan, Elena L Glassman, and Tianyi Zhang. 2021. Visualizing Examples of Deep Neural Networks at Scale. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.

[81] Han Yang, Ruimao Zhang, Xiaobao Guo, Wei Liu, Wangmeng Zuo, and Ping Luo. 2020. Towards photo-realistic virtual try-on by adaptively generating-preserving image content. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 7850–7859.

[82] Liu Yang, Weike Chen, and Lin Xu. 2019. Skin map-aided skin smoothing of images using a bilateral filter. US Patent App. 16/079,308.

[83] Qiang Yang, Yang Liu, Yong Cheng, Yan Kang, Tianjian Chen, and Han Yu. 2019. Federated learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning* 13, 3 (2019), 1–207.

[84] Manna Zelealem. 2017. Diversity in the Beauty Industry. (2017).

[85] Tingting Zhang, Ling Cao, and William YC Wang. 2017. The Impact of Virtual Try-on Image Interaction Technology on Online Shoppers' Purchase Decision. In *Proceedings of the 8th International Conference on E-Education, E-Business, E-Management and E-Learning*. 6–10.

[86] Tingting Zhang, William Yu Chung Wang, Ling Cao, and Yan Wang. 2019. The role of virtual try-on technology in online purchase decision from consumers' aspect. *Internet Research* (2019).

[87] Yunfeng Zhang, Rachel KE Bellamy, and Wendy A Kellogg. 2015. Designing information for remediating cognitive biases in decision-making. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 2211–2220.

A APPENDIX

A.1 Measurements of Product Selection Perception

The measurements of users' skincare product selection perception is shown in Table 5.

Table 5: Measurements of users' product selection perception.

	Item	Source	Cronbach's α
Website Usability	Q1. It is easy to use. Q2. It is easy to learn. Q3. I would like to recommend this website to others.	[74, 76, 78]	0.70
Website Usefulness	Q4. This website helps me select better skincare products. Q5. This website helps me select skincare products faster.	[48]	0.91
Website Credibility	Q6. The website is trustworthy. Q7. The website is professional. Q8. The website is believable. Q9. The website is expert.	[31, 64]	0.88
User Confidence in Choice	Q10. I am confident in my choices.	[53, 77, 80]	–
Website Ability of Improving User Confidence in Choice	Q11. This website improves my confidence in my choices.	[55]	–
User Satisfaction with Website	Q12. I am satisfied with this website.	[53, 55, 77]	–
Perceived Informativeness	Q13. The information needed to select suitable skincare products is easy to access. Q14. The information needed to select suitable skincare products is rich. Q15. The information is sufficient to select suitable skincare products.	[74, 76, 78]	0.89
Trust of Website	Q16. I trust the website.	[71]	–