Computers in Industry 144 (2023) 103780

Contents lists available at ScienceDirect

Computers in Industry

journal homepage: www.sciencedirect.com/journal/computers-in-industry

# On the application of extended reality technologies for the evaluation of product characteristics during the initial stages of the product development process

Almudena Palacios-Ibáñez [a,*], Raúl Navarro-Martínez [b], Joaquín Blasco-Esteban [b], Manuel Contero [a], Jorge D. Camba [c]

[a] Instituto Universitario de Investigación en Tecnología Centrada en el Ser Humano, Universitat Politècnica de València, Ciudad Politécnica de la Innovación (8B), Camí de Vera s/n, 46022 Valencia, Spain
[b] Universitat Politècnica de València, Camí de Vera, 46022 Valencia, Spain
[c] School of Engineering Technology, Purdue University, 401 Grant St, West Lafayette, IN 47907, USA

ARTICLE INFO

*Keywords:*
Virtual reality
Augmented reality
Semantic Differential
Design Representation
Product evaluation
Perception

ABSTRACT

Fast-growing global markets are forcing companies to continuously re-assess customer needs when designing new products. Product evaluation is a critical task to ensure success, but it can require significant financial and time investments. From an end-user standpoint, consumers must also evaluate multiple design options before purchasing a product, which is often a complex process, especially in online environments where traditional formats coexist with more sophisticated media. Modern extended reality technologies have become an effective tool for product assessment in professional design environments as well as a powerful mechanism for consumers during decision making activities. However, the modality used to view and evaluate the product may affect the perceptual response and thus the user's overall evaluation. In this paper, we examine the influence of visual media in product assessment using different designs of a particular product typology. We discuss two studies where a group of participants used the semantic differential technique to evaluate four chair designs displayed in three different media. In our first study, participants used simultaneous evaluation to assess the products as presented in photographs, a non-immersive environment, and an Augmented Reality (AR) experience. In the second study, participants evaluated the product separately as viewed in non-photorealistic rendering, AR, and virtual reality (VR). We used the "Aligned Rank Transform" proceedings to find differences between groups for the semantic scales, the overall evaluation, the purchasing decision, and the response confidence. Our results show that visual media influences product perception. Certain characteristics in Jordan's physio-pleasure category are particularly significant as perceptual differences are more pronounced. Immersive media can highlight some product attributes and a joint evaluation can help minimize these differences.

## 1. Introduction

Fast-growing global markets are forcing companies to continuously re-evaluate consumer needs when designing new products (Coutts et al., 2019). Consumers must decide among a wide range of functional products, where market saturation has led to an increasing supply of products with high emotional value (Aftab and Rusli, 2017).

Although the level of innovation of these products can affect consumer choice, it is not the main factor for product success. In fact, many innovative products fail when they reach the market (Marquis and Deeb, 2018; Van Kleef et al., 2005). Some researchers have suggested that success is strongly linked to product evaluation throughout the design process (Cooper, 2019).

Physical prototyping is a popular tool for evaluating design concepts. Several authors have studied how product-user interaction influences product success (Desmet et al., 2008). However, the physical prototyping process may involve significant investments in terms of time and money, with limited flexibility to modifications, and even efficient techniques such as 3D printing and rapid prototyping do not significantly reduce the amount of time invested in many cases.

* Corresponding author.
*E-mail address:* alpaib@doctor.upv.es (A. Palacios-Ibáñez).

https://doi.org/10.1016/j.compind.2022.103780
Received 22 May 2022; Received in revised form 2 September 2022; Accepted 4 September 2022
Available online 23 September 2022

The availability and affordability of Extended Reality (XR) technologies such as Virtual Reality (VR) and Augmented Reality (AR) have fueled their adoption in a number of sectors, including product development (Berni and Borgianni, 2020). New devices with improved quality, usability, ergonomics, and efficiency have made virtual prototyping an effective tool to represent products in various industrial and design applications (Berni et al., 2020). Today, XR is used throughout the design process, particularly in the early stages of development where many design variations need be produced quickly (Cecil and Kanchanapiboon, 2007).

The cost of making a design change increases dramatically as the product moves through its lifecycle (Van Kleef et al., 2005). In this regard, XR technologies can help reduce design costs by enabling engineers to study and improve the product in virtual space before it is passed on to manufacturing. XR technologies have also proven to be an effective alternative for involving final users in the design process by facilitating the collaboration between designers and users (Bruno and Muzzupappa, 2010). Finally, although the creation of a Virtual Environment (VE) may require a considerable time investment depending on the desired level of realism, XR technologies allow the real-time modification of certain product features (i.e. textures, colors or materials), which enables the exploration of a large number of design alternatives without the need for physical prototypes. In this regard, many authors have employed these techniques in studies on design evaluation (Arbeláez and Osorio-Gómez, 2018; Cascini et al., 2020; De Crescenzio et al., 2019; Song et al., 2018).

XR have also had an effect on online retail and e-commerce, whose popularity (Wang et al., 2020) has been increasing steadily in recent years, especially during the COVID-19 pandemic (Tran, 2021). Today, products are presented in online platforms in ways that go beyond traditional 2D images (Galán et al., 2021a), making the presentation medium a key differentiating factor to make a product successful. Some authors have discussed the physical barrier that exists in online platforms between user and product, which can be overcome by making attractive high-quality product presentations that allow consumers to evaluate the product accurately. In this regard, new visualization technologies have been gaining traction in recent years as mechanisms to enable richer user-product interactions. They are in high demand (Statista, 2021) and concepts such as the "metaverse" are becoming more popular.

Current technologies enable a wide range of possibilities for presenting a product, including non-immersive VEs, where content is displayed through traditional interfaces (Pleyers and Poncin, 2020), immersive VEs experienced through Head Mounted Displays (HMD's) (Jerald, 2015), or technologies such as AR (Arbeláez and Osorio-Gómez, 2018), which integrate virtual products into real environments. However, the level of fidelity and realism of the prototypes in different stages of the design process may vary depending on the specific needs (Chu et al., 2022). Modern devices can even collect physiological data such as eye-tracking, heart rate, or cognitive load, thus allowing subconscious user opinions to be translated into new design requirements in a non-invasive manner.

In this context, a critical factor in the decision-making process that occurs early in the product development process is that the evaluation provided by test users using XR technologies must be as accurate and reliable as possible. These technologies enable the creation of high-fidelity geometric representations to evaluate a product (Bordegoni, 2011), positively affecting the user's confidence and accuracy in the assessment (Hannah et al., 2012). While some aspects such as aesthetic features or visual quality can be effectively assessed using mixed realities, product features that require physical interaction still rely on physical prototypes (Bordegoni, 2011). It is generally assumed that our perceptual and emotional response to a product perceived using XR technologies is comparable to that of the physical product. However, research has shown that this is necessarily not the case (Galán et al., 2021a, 2021b; Felip et al., 2019).

Product form can also influence product perception (Achiche et al., 2014). Form plays a significant role in the aesthetics of a product and it is a critical aspect during the design process. Form also usually conveys the first impression about the product to the user. In this context, XR technologies can be leveraged to understand the influence of geometry on product perception. Products with atypical and complex geometrical shapes may be more likely to elicit perceptual differences on features related to product aesthetics when viewed in different visual media (Palacios-Ibáñez et al., 2022). Determining whether these results are generalizable to other product typologies or more traditional designs can provide valuable information to inform product design processes and decisions at the point of sale.

In addition, conjoint analysis (Eggers et al., 2022) can be used to study whether the emotional responses elicited by a product are comparable when the product is viewed separately or in context when surrounded by other designs. Generally, in a retail environment (both physical and online) different designs of the same product typology are available to the user at any given time, so a joint evaluation is a more prevalent and realistic scenario. Authors Hsee et al (Hsee et al., 1999). reported that when people evaluate products separately, relevant attributes that are difficult to evaluate are likely to be neglected in favor of attributes that are irrelevant but salient. Therefore, it should be more difficult for consumers to assess certain product attributes when evaluating one product at a time rather than when multiple products are evaluated simultaneously (Christopoulos et al., 2011). Since new product designs are often evaluated individually throughout the design process, it could be argued that these assessments may not be entirely accurate compared to joint evaluations in VEs, which can help minimize perceptual differences.

The present study contributes to our understanding of how XR technologies can affect the various dimensions of the perceptual space linked to a product using different evaluation methods. The study examines how XR technologies affect the user's emotional response during product evaluation, assessing whether these technologies can make the design process less time-consuming and more cost-effective. Furthermore, we investigate how the evaluation format (individual or joint) affects product assessment, and whether providing several concept options to the users can yield more accurate evaluations. We applied the semantic differential technique (Osgood et al., 1957) to evaluate the characteristics of two sets of chairs with very different characteristics.

## 2. Background

The influence of the presentation medium on user perception has been studied in different fields (Chu et al., 2022; Bordegoni, 2011). The effect of visual presentation media on product perception was first studied by Söderman (2005) who compared assessments of cars in a non-immersive VR environment and 2D sketches to evaluations of physical cars. His results showed that the perception of certain product features was not affected by the display medium, which the author attributed to the user's prior knowledge of the product.

Artacho-Ramírez et al. (2008). studied the perceptual variations of two loudspeaker designs in five different media, concluding that the type of media significantly influenced product perception. Other studies incorporated more sophisticated technologies (Kato, 2019). For example, Galán et al. (2021a). used household products that included passive haptics in which physical objects were synchronized with virtual counterparts to allow users to physically feel some of the virtual objects they saw and interacted with. The authors found perceptual differences due to the variation of the medium, especially for those in Jordan's category of physical pleasure (Jordan, 2002).

In the realm of AR, Ray and Choi (2017) investigated how AR affects product assessment and Agost et al. (2021) reported a study in which a lamp and a sideboard were evaluated using 2D images, 360-degree visualizations, AR, and VR. Perceptual differences showed that AR is an effective medium for displaying large objects, whereas 2D images are

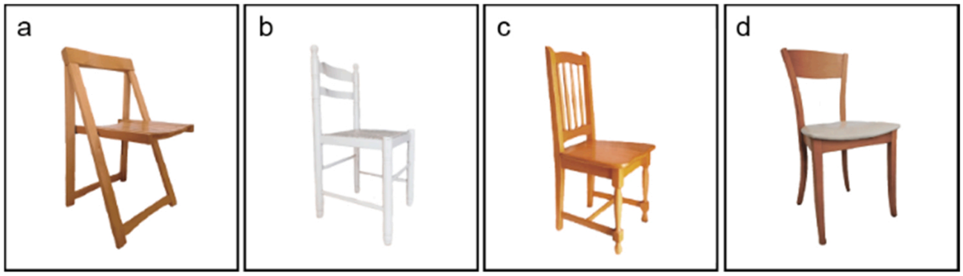


**Fig. 1.** Photographs of the chairs used in the first study.

effective for smaller objects. 360-degree visualizations are suitable for both types of products.

Few authors have employed conjoint product evaluation in their studies on perceptual analysis with different means of evaluation. Chuang et al. (2001). used this methodology to jointly evaluate 26 cell phones using the semantic differential technique to examine the relationship between users' preferences and the geometric design elements of the products. Some researchers have also studied the perceptual differences that arise when changing the presentation format of a product, but observations have only been made on a case-by-case basis. Furthermore, consumers often evaluate different designs of the same product typology before making a purchasing decision. Therefore, studying how a user's perception of a single product varies individually when changing the display medium may not provide results that are applicable to an everyday situation. Instead, simultaneous evaluation methods may provide more accurate insights.

Although new visualization methods are gradually reducing the physical barriers between user and product in virtual platforms, it is unclear the extent to which the user's perception of the product is influenced by the presentation media. In this paper, we present two studies with different chair designs as stimuli. In our first study, users evaluated four common chairs through 2D photographs of the product (IMG), a non-immersive environment (3D), and AR using a conjoint evaluation method. In the second study, we used four concept chair designs submitted to the 2014 Annual Wilsonart Student Chair Design Competition for the evaluation (the chair designs can be viewed at http://www.blogtour.co/wilsonart-does-design-proud-student-chair-design-competition/). In this case, product assessments were done individually using non-photorealistic renderings (NPR), AR, and VR.

## 3. Hypotheses

We postulated the following hypotheses: The presentation media influences the user's perception of the product (H0.1); The presentation media influences the user evaluation of semantic scales independently of their classification in Jordan's categories (H0.2); The overall evaluation of the product (H0.3), the confidence in the user's response (H0.4) and the purchase decision (H0.5) is influenced by the presentation medium; A joint evaluation can minimize perceptual differences for the semantic scales between visual media (H1.1, hypotheses postulated only for the first study).

## 4. Materials and methods

### *4.1. Case study*

To validate the previous hypotheses, we designed two experimental studies. The main purpose was to examine the interaction between product aesthetic features and modes of representation. To draw robust conclusions, we selected different designs of the same product typology (i.e. a chair) with notable aesthetic differences. Each set of chairs was studied separately. Because of their aesthetic differences, various means of presentation were selected to facilitate the understanding of each product.

In our first study, a group of 40 participants evaluated four common everyday chairs as shown in Fig. 1.

Participants ranked the chairs according to eight semantic differential scales in three different means of presentation:

- Photographs of the product (Fig. 1) taken from multiple points of view and on a white background to avoid any interferences with external stimuli. Pictures were displayed using a computer screen, and the participants were allowed to use the mouse to zoom in and out if needed, as well as the arrows keys in the keyboard to switch pictures.
- A non-immersive environment (Fig. 2) where virtual products were placed at the center of a virtual room. To minimize the impact of the environment on the product's evaluation or attention, the environment consisted of a simple shape and large room with neutral colors. This setting was displayed on a computer monitor, and no interaction with the virtual product was allowed, but the user was able to navigate the space by using their mouse and different keys on the keyboard to look at the product from any angle.
- Augmented reality (Fig. 3), where the virtual products were placed in the real world. Since there was limited control over the surrounding environment in this setting, the participant was asked to display the

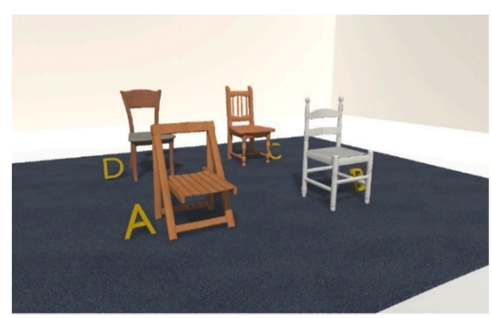


**Fig. 2.** Interactive 3D environment used in the first study.


**Fig. 3.** Virtual products placed in a real environment (AR) for the first study.

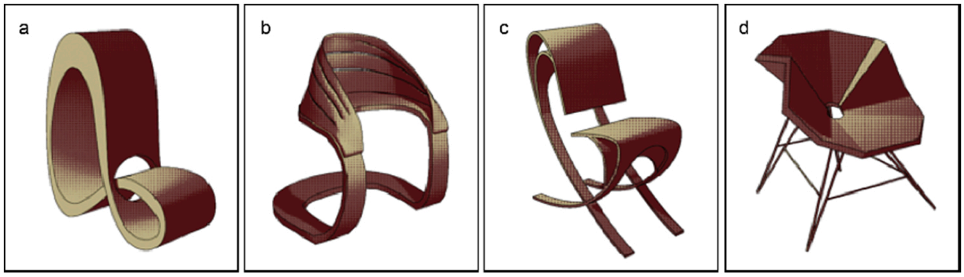


**Fig. 4.** NPR of the chairs used in the second study (Credit: Jenny Trieu (A), Abizer Raja (B), Arturo Barrera (C), and Carrah Kaijser (D)).

**Fig. 5.** Virtual product placed in a real environment (AR) for the second study.

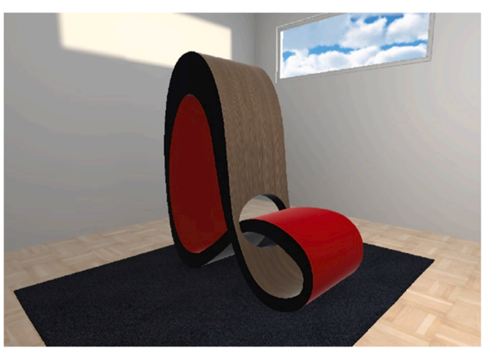

**Fig. 6.** Virtual product placed in a virtual environment (VR) for the second study.

objects in a common usage environment with a clear area for the models to have enough space. The user was allowed to move around the objects to view them from any angle. As in the previous environment, interaction with the products was not allowed.

The means of representation selected for the first study were considered to provide enough information to the participants, as they were all common designs.

In the second study, 32 participants evaluated four unconventional chairs individually using a seven-level semantic scale. The media used included:

- A set of 2D images synthesized through non-photorealistic rendering techniques (Fig. 4). The images were generated from different points of view and on a white background to avoid any interferences with external stimuli during the product assessment. As in the first study, these pictures were displayed on a computer screen, and the participant was allowed to use the mouse to zoom in and out if needed, and the arrow keys in the keyboard to switch pictures.
- AR (Fig. 5), where the virtual product was placed in a real environment with similar characteristics to the first study. Interaction with the virtual product was not allowed, but the user was allowed to move freely.
- VR (Fig. 6), where the virtual product was placed in a VE experienced via an HMD. Product interaction was not allowed, but the user was allowed to move around the object and view it from different points of view.

We considered VR for the second study as a medium that could provide valuable information to compensate for the fact that participants may not have relevant previous knowledge about these chairs and thus not be as familiar with the characteristics of the products.

For both studies, participants were asked to rate how much they liked the product as well as their level of confidence in their responses using a 5-point Likert scale (where 1 represented “Dislike” and “no confidence,” and 5 represented “Like” and “total confidence”) and indicate their purchasing decision with a “Yes” or “No” answer.

### *4.2. Semantic differential scales for product evaluation*

For our two studies, a set of semantic differential scales (Osgood et al., 1957) composed of eight bipolar pairs of adjectives serving as chair descriptors were used for product evaluation. This rating scale is a common method of product assessment (Desmet, 2002) that does not force respondents to discriminate between items and allows them to state that several items are of similar importance (Flynn and Marley, 1992). To determine the semantic differential scales for our experiment, we compiled a list of adjectives from similar studies that uses the same (or similar) product as stimulus (Galán et al., 2021b; Felip et al., 2019). The final bipolar pairs of adjectives are shown in Tables 1 and 2 and are classified according to the four categories described by Jordan (2002). Due to the significant geometrical differences between the selected stimuli (a classical and homogeneous design in the first case study vs. an atypical design in the second), a decision was made to generate a slightly

**Table 1**
List of semantic scales used in study 1 classified by Jordan’s pleasure categories.

| Physio-pleasure | Psycho-pleasure | Socio-pleasure | Ideo-pleasure |
|---|---|---|---|
| Comfortable – Uncomfortable | Complex - Simple | Classic – Modern | Industrial - Handmade |
| Well proportioned – Unproportioned | Minimalistic – Overelaborated | Elegant – Conventional | Fun - Serious |

**Table 2**
List of semantic scales used in study 2 classified by Jordan's pleasure categories.

| Physio-pleasure | Psycho-pleasure | Socio-pleasure | Ideo-pleasure |
|---|---|---|---|
| Comfortable - Uncomfortable | Complex – Simple | Classic – Modern | Industrial – Handmade |
| Light – Heavy | Practical – Impractical | Attractive – Unattractive | Fun – Serious |

different semantic differential. In the first case study, a rank method was used to evaluate each product. For the second study, we opted for seven-level scales.

### 4.3. Materials

All the chairs in our studies were modeled in Blender 2.93.0. Textures had an image size of 512px and were acquired from Adobe Substance 3D Assets. The VR environment and 3D were designed using Unity 2019.4.14f1 with baked lights. The AR environment was created using the online resource Clon Digital (https://clondigital.es), which enables the integration of 3D models in a real environment without the need to develop a custom application.

To correctly visualize the non-immersive environment, we used Simmer (https://simmer.io), an online repository for Unity WebGL games. The VR environment was displayed using the Oculus Quest 2 HMD, a standalone immersive VR device with a Single Fast-Switch LCD of 1832 × 1920 pixels per eye and a refresh rate of 72 Hz. For the AR environment in the first study, we used a wide range of smartphones with a minimum API level of 7 for Android devices, and iOS 11 version and an A9 processor as minimum specifications for iOS devices. For the second study, we used a OnePlus 7 T, a smartphone with a screen size of 6,55in and a 1080 × 2400 screen resolution.

Due to COVID-19 restrictions, some interviews were conducted online for the first experiment via the Discord platform (instant messaging service for voice chat, video, and text chat). IBM SPSS Statistics 22 and Microsoft Excel were used for the inferential statistical analysis.

### 4.4. Sample

An a priori power analysis with G*Power (Faul et al., 2007) was conducted to estimate the minimum sample size, assuming an ANOVA with repeated measures with the following input parameters: effect size: 0.25, $\alpha = .05$, $(1\text{-}\beta)= .80$ and 1 group. Our results estimated a total sample size of 28. To guarantee a power of.80, a total of 40 volunteers participated in the first experiment (25 men and 15 women, average age: 32 years old). The experiment was conducted both in person and online (62,50 % of the participants in person, and 37.50 % online). Before the experiment, users were asked to rate their experience with AR using a four-point Likert scale from 0 to 3 (0 = no experience, and 3 = significant experience). A total of 42.50 % of the participants had no previous experience with AR, 37.50 % rated their experience as limited, 17.50 % stated having a lot of experience with AR, and 2.50 % rated their experience as significant.

A total of 32 participants took part in the second study: 25 % were male and 75 % female with an average age of 24,88 years old. 50 % of the male participants and 53,1 % of the female participants had no previous experience with VR and AR, 34.4 % of males and 28.1 % of females rated their experience as limited with VR and AR, 15.6 % claimed to have vast experience with VR and AR, and 3.1 % rated their experience as significant with AR.

### 4.5. Experimental protocol

For both studies, all volunteers were over 18 years old. Before starting the experiment, verbal consent was obtained from each participant as well as basic demographics information (gender and age), data about the user experience in AR (for both studies) and VR (only for the second). Participants experienced the three experimental conditions (the viewing media) in a random sequence to minimize the effect of the presentation order. No interaction with the product was allowed in any of the visual media.

In the first study, participants were exposed to the four chairs simultaneously in each media and asked to rank them using the eight bipolar pairs in Table 1. They also evaluated how much they liked each chair and rated the level of confidence in their responses using a 5-point Likert scale. Finally, they were asked to make a purchasing decision ("Yes" or "No"). Each participant spent an average of 22 min per interview.

The second experiment was conducted entirely in person and involved the use of an HMD. In this case, participants were exposed to each chair individually in each media and asked to evaluate them using a eight seven-level semantic scales. Each participant completed a total of 12 evaluations. They also evaluated how much they liked each chair and rated the level of confidence in their responses using a 5-point Likert scale. Finally, they were asked to make a purchasing decision ("Yes" or "No"). The experiment took an average duration of 45 min per interview.

## 5. Results

In order to obtain more robust and reliable results, participants who experienced difficulties interpreting and applying the semantic scales were identied as outliers. As an additional factor, we decided to eliminate only those data points that appeared as outliers in at least 3 chairs. As a result, 5 data points were deleted for the first study.

### 5.1. First study results

Chairs were ranked according each bipolar pair presented on Table 1. The score obtained by each chair was equivalent to its position on the bipolar pair rank. Each end of the scale was associated to a particular adjective of a pair, so that a score closer to 1 indicated a greater correspondence with that adjective, and a score closer to 4 indicated a greater correspondence with the opposite adjective.

Four different data sets were obtained: the semantic scales, the overall evaluation, the purchase decision, and finally, the level of confidence in the response for each media. A normality test was performed on each data set to select the appropriate statistical test. As the sample size was less than 50 participants, we used a Shapiro-Wilks's normality test (significance level of.05). Results showed that the data was not normally distributed, so parametric tests were unsuitable.

We applied the Aligned Rank transform (ART) procedure (Higgins et al., 1990) as it provides a powerful and robust nonparametric alternative to other traditional techniques (Mansouri et al., 2004). It relies on a preprocessing step that "aligns" data before applying averaged ranks. After this step, common ANOVA procedures can be applied. The descriptive statistics for our four data sets are shown in Tables 3 to 5 and semantic scales stacked bar charts are shown in Fig. 7. The adjective in bold corresponds to a score of 4, whereas the adjective in italics (Ozok and Komlodi, 2009) corresponds to a score of 1.

Next, differences between the display techniques were analyzed. One factor repeated measures ANOVAs and post-hoc tests with Bonferroni correction were performed for the semantic scales (Tables 6 and 7). Although the p-value of the bipolar pair "Well proportioned - Unproportioned" was.036 for Chair B-1, post-hoc tests did not find significant differences in the pairwise comparisons.

We also performed a one-factor repeated measures ANOVA for the

**Table 3**
Descriptive statistics for semantic scales (study 1).

| Semantic scales | | Chair A-1 | | | Chair B-1 | | | Chair C-1 | | | Chair D-1 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | IMG | 3D | AR | IMG | 3D | AR | IMG | 3D | AR | IMG | 3D | AR |
| | Mean | **1.63** | 1.57 | *1.49* | *2.31* | 2.69 | **2.89** | 2.31 | **2.37** | *2.29* | 3.74 | **3.37** | *3.34* |
| **Comfortable** – *Uncomfortable* | Med. | 1.00 | 1.00 | 1.00 | 2.00 | 3.00 | 3.00 | 2.00 | 2.00 | 2.00 | 4.00 | 4.00 | 4.00 |
| | SD | .81 | .79 | .66 | .83 | .87 | .99 | .99 | 1.14 | 1.02 | .61 | .88 | .84 |
| | Mean | **1.60** | *1.31* | 1.40 | *2.51* | 2.89 | **3.00** | 3.34 | **3.40** | *3.31* | **2.54** | 2.40 | *2.34* |
| **Well proportionated** - *Unproportionated* | Med. | 1.00 | 1.00 | 1.00 | 3.00 | 3.00 | 3.00 | 4.00 | 4.00 | 4.00 | 3.00 | 3.00 | 3.00 |
| | SD | .78 | .63 | .55 | 1.04 | .90 | .80 | .80 | .81 | .83 | 1.12 | .91 | 1.11 |
| | Mean | *2.86* | 2.89 | **2.97** | **2.89** | *2.57* | 2.69 | **1.63** | 1.49 | *1.43* | *2.63* | 3.06 | **3.37** |
| **Simple** – *Complex* | Med. | 3.00 | 3.00 | 3.00 | 3.00 | 2.00 | 3.00 | 2.00 | 1.00 | 1.00 | 3.00 | 3.00 | 4.00 |
| | SD | 1.22 | 1.13 | 1.12 | .99 | .95 | .87 | 1.02 | .66 | .65 | 1.14 | 1.00 | .88 |
| | Mean | **3.34** | *3.14* | 3.31 | *2.46* | **2.57** | **2.57** | 1.57 | **1.63** | *1.54* | 2.60 | **2.66** | *2.57* |
| **Minimalistic** – *Overelaborated* | Med. | 4.00 | 3.00 | 3.00 | 2.00 | 2.00 | 3.00 | 1.00 | 1.00 | 1.00 | 3.00 | 3.00 | 2.00 |
| | SD | .90 | .97 | .80 | .98 | 1.04 | .95 | .85 | .94 | .92 | 1.04 | 1.00 | 1.07 |
| | Mean | *2.29* | **2.49** | *2.29* | **3.46** | *3.37* | 3.43 | 1.94 | *1.89* | **1.97** | *2.31* | **2.26** | *2.31* |
| **Classic** – *Modern* | Med. | 2.00 | 2.00 | 2.00 | 4.00 | 4.00 | 4.00 | 1.00 | 1.00 | 2.00 | 2.00 | 2.00 | 2.00 |
| | SD | .86 | .98 | 1.10 | .70 | .84 | .78 | 1.19 | 1.11 | 1.01 | 1.08 | 1.01 | 1.02 |
| | Mean | 1.54 | *1.46* | **1.60** | *2.11* | **2.31** | 2.14 | *3.29* | **3.34** | *3.29* | **3.09** | *2.89* | 3.06 |
| **Elegant** – *Conventional* | Med. | 1.00 | 1.00 | 1.00 | 2.00 | 2.00 | 2.00 | 3.00 | 4.00 | 4.00 | 3.00 | 3.00 | 3.00 |
| | SD | .85 | .74 | .85 | .99 | .90 | .88 | .79 | .91 | 1.02 | .85 | .96 | .84 |
| | Mean | 3.31 | **3.40** | *3.29* | *1.49* | **1.57** | 1.54 | 2.57 | **2.63** | *2.54* | **2.63** | *2.40* | **2.63** |
| **Industrial** – *Handmade* | Med. | 4.00 | 4.00 | 4.00 | 1.00 | 1.00 | 1.00 | 2.00 | 3.00 | 2.00 | 3.00 | 3.00 | 3.00 |
| | SD | .90 | 1.04 | 1.07 | .78 | .66 | .82 | .98 | .84 | .980 | 1.00 | 1.09 | .88 |
| | Mean | **2.77** | *2.69* | **2.77** | **2.77** | *2.49* | 2.54 | *1.91* | **2.11** | 2.09 | 2.54 | **2.63** | *2.51* |
| **Serious** – *Fun* | Med. | 3.00 | 3.00 | 3.00 | 3.00 | 2.00 | 3.00 | 1.00 | 2.00 | 2.00 | 2.00 | 3.00 | 2.00 |
| | SD | 1.00 | .93 | 1.06 | 1.00 | 1.04 | 1.04 | 1.15 | 1.21 | 1.12 | 1.15 | 1.24 | 1.20 |

Highest values and corresponding adjective in bold. Lowest values and corresponding adjective in italics.

**Table 4**
Descriptive statistics for overall evaluation and purchasing decision (study 1).

| | | Chair A-1 | | | Chair B-1 | | | Chair C-1 | | | Chair D-1 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | IMG | 3D | AR | IMG | 3D | AR | IMG | 3D | AR | IMG | 3D | AR |
| | Mean | 2.31 | **2.34** | *2.14* | 2.69 | **3.03** | *2.57* | *3.49* | **3.51** | **3.51** | **3.37** | 3.06 | *2.74* |
| Like/Dislike | Med. | 2.00 | 2.00 | 2.00 | 3.00 | 3.00 | 3.00 | 4.00 | 4.00 | 4.00 | 3.00 | 3.00 | 3.00 |
| | SD | .72 | .85 | .85 | .87 | .82 | .78 | .74 | .85 | .70 | .77 | .87 | .95 |
| | Mean | **.06** | **.06** | *.03* | *.06* | **.11** | .09 | *.26* | *.26* | **.29** | *.17* | **.20** | *.17* |
| Purchasing decision | Med. | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 |
| | SD | .236 | .236 | .169 | .236 | .323 | .284 | .443 | .443 | .458 | .382 | .406 | .382 |

Highest values in bold. Lowest values in italics.

**Table 5**
Descriptive statistics for the response confidence (study 1).

| | | IMG | 3D | AR |
|---|---|---|---|---|
| | Mean | **3.74** | 3.63 | *3.43* |
| Response confidence | Med. | 4.00 | 4.00 | 3.00 |
| | SD | .61 | .65 | .66 |

Highest values in bold. Lowest values in italics.

overall evaluation (Table 8), and a post-hoc analysis for Chairs B-1 and D-1 are shown in Table 9. Cochran's Q test was performed for the purchase decision (Table 10), but no significant differences were found between means. Finally, no statistically significant differences were found for the response confidence, F(2, 68) = 2.474, p = .092.

### 5.2. Second study results

Descriptive statistics for our second study are shown in Tables 11–13. A value closer to − 3 represents a closer correspondence with the adjective to the left, and a value closer to 3 indicates a closer correspondence with the adjective to the right. Stacked bar charts for the semantic scales are shown in Fig. 8. A Shapiro-Wilk's normality test (significance level of.05) revealed that the data was not normally distributed, so parametric tests proved unsuitable. Therefore, the ART procedure was applied once again.

We performed one-factor repeated measures ANOVA and Post-hoc tests (Bonferroni adjustment was applied) for each data set. For the purchasing decision, we performed Cochran's Q and McNemar tests. The results are shown in Tables 14 to 16. Although the p-value of the bipolar pair "Modern-Classic" was.046 for Chair B-2, post-hoc tests did not find significant (Tables 17, 18) differences in the pairwise comparison (Table 15).

Finally, the response confidence, statistically significant differences were found for Chairs A-2 (F(2, 62)= 8.56, p = .001) and D-2 (D-2 F (2,62)= 3.73, p.030). Post-hoc tests are shown in Table 19.

## 6. Discussion

In this paper, we examined the influence of the presentation medium on the perception of a set of chairs using different evaluation methods. We conducted two different studies: in the first case, four chairs with a classic and homogeneous design were selected to be viewed and evaluated simultaneously using photographs of the actual product, a non-immersive virtual environment, and AR. In the second case study, a more diverse set of four chairs with an atypical design were presented to participants and evaluated individually using a set of 2D images synthesized through non-photorealistic rendering techniques, AR, and VR. In both studies, participants used the semantic differential technique to evaluate the products. In addition, participants were asked to provide an overall evaluation of each chair ("Like/Dislike"), a purchase decision, and a rating of the level of confidence in their responses for each visual medium.

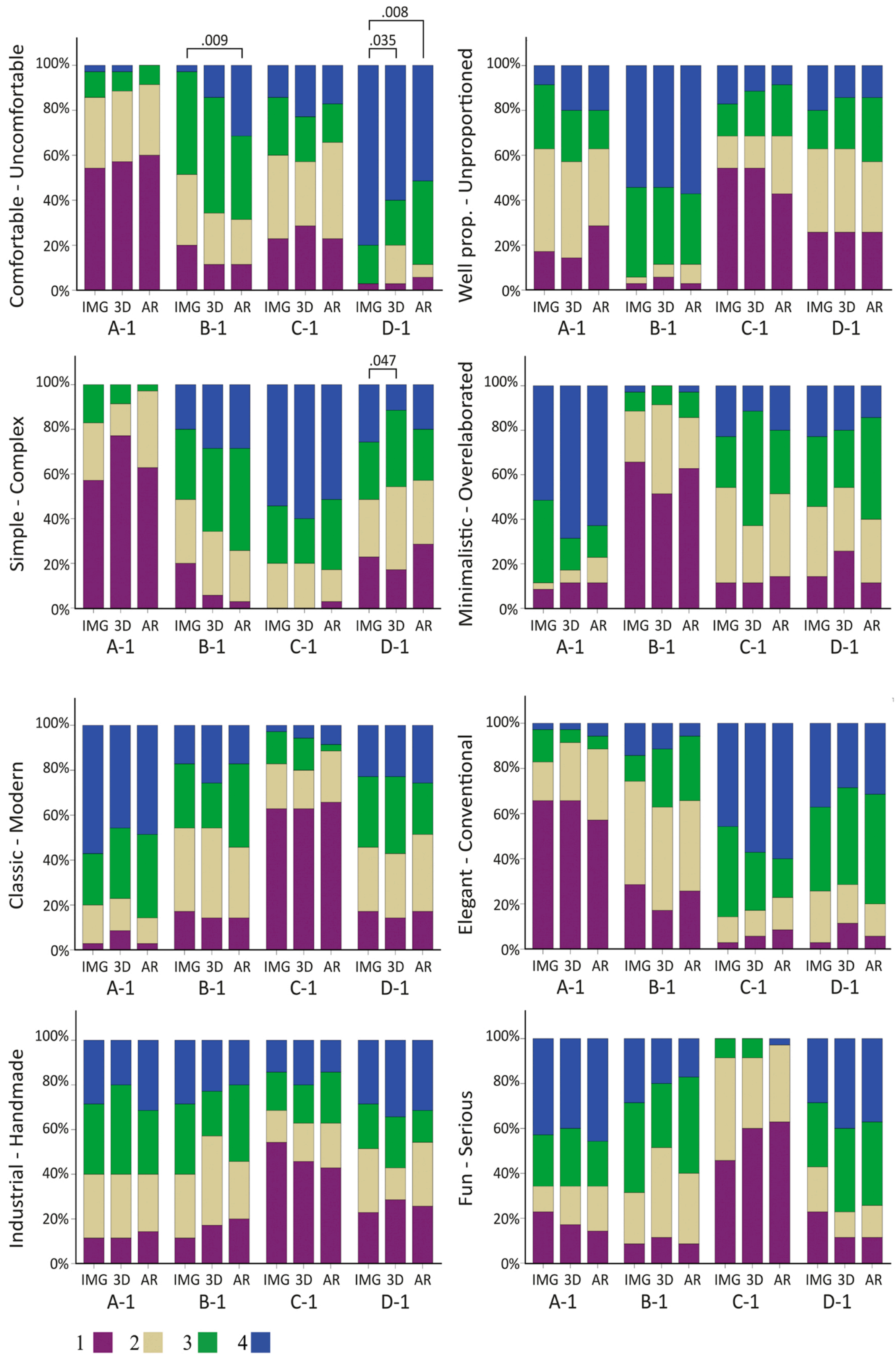


**Fig. 7.** Stacked bar charts for the semantic differential scales (study 1).

**Table 6**
Repeated measures ANOVA for the semantic scales (study 1).

| | Semantic scales | df | Chair A-1 | | Chair B-1 | | Chair C-1 | | Chair D-1 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | F | Sig. | F | Sig. | F | Sig. | F | Sig. |
| PHYSIO | Comfortable – Uncomfortable | 2 | 2.166 | p = .122 | 6.567 | **p = .002** | .069 | p = .933 | 6.088 | **p = .004** |
| | Well proportionated - Unproportionated | | 2.166 | p = .122 | 3.502 | **p = .036** | .175 | p = .840 | .578 | p = .564 |
| PSYCHO | Simple – Complex | | .262 | p = .770 | 1.732 | p = .185 | 2.764 | p = .070 | 3.992 | **p = .023** |
| | Minimalistic – Overelaborated | | .851 | p = .432 | .315 | p = .731 | .194 | p = .824 | .150 | p = .861 |
| SOCIO | Classic – Modern | | .962 | p = .387 | .051 | p = .950 | .486 | p = .617 | .086 | p = .918 |
| | Elegant – Conventional | | .626 | p = .538 | 2.123 | p = .128 | .550 | p = .580 | .594 | p = .555 |
| IDEO | Industrial – Handmade | | 1.559 | p = .218 | 1.211 | p = .304 | .317 | p = .729 | 1.482 | p = .235 |
| | Serious – Fun | | .147 | p = .864 | 1.180 | p = .313 | .726 | p = .488 | .185 | p = .831 |

p-values less than.05 are shown in bold.

**Table 7**
Post-hoc tests for the semantic scales (chair B-1 and D-1).

| Semantic scales | Means | Chair B-1 Sig. | Chair D-1 Sig. |
|---|---|---|---|
| Comfortable – Uncomfortable | IMG – 3D | p = .054 | **p = .035** |
| | IMG - AR | **p = .009** | **p = .008** |
| | 3D - AR | p = .563 | p = 1.000 |
| Well prop. – Unproportioned | IMG – 3D | p = .175 | |
| | IMG - AR | p = .085 | |
| | 3D - AR | p = 1.000 | |
| Simple – Complex | IMG – 3D | | **p = .047** |
| | IMG - AR | | p = .169 |
| | 3D - AR | | p = 1.000 |

p-values less than.05 are shown in bold.

Our results show that the purchase decision is not influenced by the visual medium. Differences between media for this dataset were found only for Chair A-2 (H0.5 is rejected). Therefore, it could be argued that although access to information is critical to make a purchase decision (O'Keefe and McEachern, 1998), the medium used to present this information may not be a determining factor. These results agree with other studies that concluded that 2D media may offer sufficient information to reliable assess a product (Ant Ozok and Komlodi, 2009).

In our first hypothesis (H0.1), we speculated that the presentation media could influence the user's perception of the products. In the first case study, perceptual differences were found for certain bipolar pair of adjectives for Chairs B-1 and D-1 (Table 6). Differences were also found in all chairs for certain bipolar pairs of adjectives in our second study (Table 14). These results confirm H0.1 for both studies, which agree with the results obtained by other authors (Palacios-Ibáñez et al., 2022) and contributes to expand the scope of product typologies. Similar to Artacho-Ramírez et al., who studied the influence of the graphical representation in the evaluation of different models of a loudspeaker (Artacho-Ramírez et al., 2008), and Agost et al., who reached similar conclusions when assessing two different type of furniture (a sideboard and a lamp) (Agost et al., 2021), our study confirms that these results can be extrapolated to other types of products.

In H0.2, we questioned whether the presentation media influences the user evaluation of semantic scales independently of their classification in Jordan's categories. In the case of Chair B-1, differences were found for the bipolar pair "Comfortable - Uncomfortable" (physiological pleasure category), and for Chair D-1, differences were found for "Comfortable - Uncomfortable" and "Simple – Complex" (psychological pleasure category). In this case, the physio-pleasure category was the most affected by the change of visual medium, which aligns with other authors who confirmed the importance of haptics for the evaluation of these characteristics (Galán et al., 2021a). The absence of touch in our study may have had a negative influence on the evaluation of certain product attributes.

Although some adjectives from the physio-pleasure category were affected by the change of medium in the second case study, the socio-pleasure category (closely related to product aesthetics) was the most influenced. These results may have been affected by the geometric characteristics of the chairs, as these products had a higher aesthetic value. It has been argued that the aesthetic elements of a product's shape can influence user perceptions (Achiche et al., 2014) so a less typical design of a chair may have influenced perception regardless of the medium in which it was presented. Our results align with those obtained by (Palacios-Ibáñez et al. 2022)., who demonstrated that the sociological pleasure category was the most affected by the change of media for the case of products with high aesthetic value (i.e., coffee makers). For our study, we also used stimuli with high aesthetic value, but significantly different features (i.e., chairs). Differences were mostly found between IMG – AR and IMG – 3D, but it is important to emphasize that different representation methods do not necessarily have the same interaction capabilities. More sophisticated media provide higher levels of interaction, so it can be expected that these perceptual differences stem from this, as demonstrated by Ant Ozok and Komlodi (2009).

According to the descriptive statistics for the semantic scales in our second study (Table 11), it is important to highlight how all four chairs appeared heavier to participants in the more immersive media (Chairs B-2 and C-2 when presented in AR, and Chairs A-2 and D-2 when presented in VR), whereas three of the four chairs (A-2, B-2, and D-2) appeared lighter when viewed by participants in NPR. This result could be attributed to the fact that the chairs were displayed in their actual sizes in AR and VR compared to flat images shown on a 2D screen (with dimensional limitations) which may have made the product appear

**Table 9**
Post-hoc tests for the overall evaluation (chair B-1 and D-1).

| | Means | Chair B-1 Sig. | Chair D-1 Sig. |
|---|---|---|---|
| Like/Dislike | IMG – 3D | p = .123 | p = .096 |
| | IMG - AR | p = .622 | **p < .001** |
| | 3D - AR | **p = .011** | p = 0.64 |

p-values less than.05 are shown in bold.

**Table 8**
Repeated measures ANOVA for the overall evaluation (study 1).

| | df | Chair A-1 | | Chair B-1 | | Chair C-1 | | Chair D-1 | |
|---|---|---|---|---|---|---|---|---|---|
| | | F | Sig. | F | Sig. | F | Sig. | F | Sig. |
| Overall evaluation | 2 | 1.486 | p = .234 | 5.940 | **p = .004** | 0.034 | p = .967 | 11.180 | **p < .001** |

p-values less than.05 are shown in bold.

**Table 10**
Cochran's Q test for the purchasing decision (study 1).

| | df | Chair A-1 | | Chair B-1 | | Chair C-1 | | Chair D-1 | |
|---|---|---|---|---|---|---|---|---|---|
| | | Q | Sig. | Q | Sig. | Q | Sig. | Q | Sig. |
| Purchase decision | 2 | 2.000 | p = .368 | 1.500 | p = .472 | .290 | p = .867 | .500 | p = .779 |

smaller (as in (Galán et al., 2021a)), and thus lighter. In addition, the NPR medium did not accurately represent the actual material of the chair, so some participants may have perceived the product as heavier in a more interactive medium. Our analyses confirm H0.2 for both studies, which agree with the studies discussed earlier.

We also speculated that, for both studies, the overall product evaluation and response confidence could be influenced by the presentation medium (H0.3 and H0.4). For the first study, our results showed that AR was the medium in which the product was less liked, and where the user felt less confident about their response, followed by 3D and IMG. Although we expected the opposite effect for the response confidence (as 3D representations provide more information to the user), our results could be explained by the user's limited experience with technology as well as the online interview conducted with some users, which could lead to lower levels of confidence when performing evaluations (42.5 % had no previous experience with AR, and 37.5 % rated their experience as limited). The lack of experience plus the pressure of taking part of an online study may have led some users to experience technology related

**Table 11**
Descriptive statistics for the semantic scales (study 2).

| Semantic scales | | Chair A-2 | | | Chair B-2 | | | Chair C-2 | | | Chair D-2 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | NPR | AR | VR | NPR | AR | VR | NPR | AR | VR | NPR | AR | VR |
| | Mean | *1.41* | 1.97 | **2.31** | *-1.63* | **-0.91** | -1.47 | -1.06 | **-0.81** | *-1.12* | *-0.5* | -0.34 | **-0.22** |
| *Light* – **Heavy** | Med. | 2.00 | 2.00 | 2.50 | -2.00 | -2.00 | -2.00 | -1.00 | -1.50 | -2.00 | -1.00 | -1.00 | 0.00 |
| | SD | 1.54 | 1.45 | 0.78 | 1.21 | 1.92 | 1.63 | 1.60 | 1.82 | 1.45 | 1.70 | 1.91 | 1.91 |
| | Mean | **1.19** | 0.25 | *0.03* | 1.09 | *1.00* | **1.13** | **-1.34** | *-1.38* | *-1.38* | 1.69 | **1.75** | *1.66* |
| *Comfortable* – **Uncomfortable** | Med. | 1.00 | 1.00 | 0.50 | 2.00 | 1.00 | 2.00 | -2.00 | -2.00 | -2.00 | 2.00 | 3.00 | 3.00 |
| | SD | 1.51 | 1.76 | 1.91 | 1.87 | 1.88 | 1.93 | 1.47 | 1.50 | 1.74 | 1.51 | 1.81 | 2.06 |
| | Mean | 1.63 | **1.69** | *1.38* | 0.00 | *-0.16* | **0.25** | -1.06 | **-0.87** | *-1.16* | *0.88* | 1.16 | **1.28** |
| *Practical* – **Impractical** | Med. | 2.00 | 2.00 | 2.00 | -0.50 | -0.50 | 1.00 | -2.00 | -1.00 | -2.00 | 1.00 | 2.00 | 2.00 |
| | SD | 1.36 | 1.65 | 1.75 | 1.80 | 1.94 | 2.00 | 1.62 | 1.64 | 1.65 | 1.74 | 1.89 | 1.85 |
| | Mean | *1.34* | 1.56 | **1.78** | *-0.22* | **0.13** | -0.16 | 0.66 | **0.75** | *0.19* | *0.09* | **0.91** | 0.69 |
| *Simple* - **Complex** | Med. | 2.00 | 2.00 | 2.00 | -0.50 | 0.00 | 0.00 | 1.00 | 1.00 | 0.50 | 0.50 | 1.00 | 1.00 |
| | SD | 1.45 | 1.58 | 1.21 | 1.84 | 1.81 | 1.72 | 1.54 | 1.70 | 1.42 | 1.92 | 1.63 | 1.65 |
| | Mean | **-2.38** | *-2.72* | -2.63 | **-0.94** | *-1.63* | -1.31 | **-1.37** | *-1.87* | -1.50 | **-2.00** | -2.28 | *-2.47* |
| *Modern* – **Classic** | Med. | -2.50 | -3.00 | -3.00 | -1.00 | -2.00 | -2.00 | -2.00 | -2.00 | -1.50 | -2.00 | -2.00 | -3.00 |
| | SD | 0.75 | 0.52 | 0.70 | 1.54 | 1.21 | 1.38 | 1.50 | 1.24 | 1.27 | 0.95 | 0.81 | 0.76 |
| | Mean | **-0.22** | -1.09 | *-1.41* | 0.25 | *-0.16* | **0.31** | -1.66 | **-1.62** | *-1.75* | 0.00 | *-0.19* | **0.16** |
| *Attractive* – **Unattractive** | Med. | -1.00 | -1.50 | -2.00 | -1.00 | -1.00 | 0.00 | -2.00 | -2.00 | -2.00 | 0.00 | -1.00 | 0.50 |
| | SD | 1.79 | 1.69 | 1.54 | 1.74 | 1.83 | 1.59 | 1.28 | 1.38 | 1.27 | 1.88 | 2.12 | 2.16 |
| | Mean | -1.69 | *-1.91* | **-1.62** | *-0.87* | **-0.69** | -0.72 | *-1.44* | -1.41 | **-1.38** | **-0.62** | *-1.06* | -0.69 |
| *Fun* – **Serious** | Med. | -2.00 | -2.00 | -2.00 | -1.00 | -1.00 | -1.00 | -2.00 | -2.00 | -2.00 | -1.00 | -1.00 | -1.00 |
| | SD | 1.09 | 1.17 | 1.31 | 1.07 | 1.35 | 1.35 | 1.43 | 1.60 | 1.60 | 1.56 | 1.46 | 1.65 |
| | Mean | **0.47** | *0.28* | *0.28* | **0.31** | *-0.75* | -0.16 | 0.41 | **0.5** | *0.13* | *0.88* | **1.13** | 1.06 |
| *Handmade* – **Industrial** | Med. | 1.00 | 1.00 | 1.00 | 1.00 | -1.00 | -1.00 | 1.00 | 1.00 | 0.00 | 2.00 | 2.00 | 1.50 |
| | SD | 2.29 | 2.14 | 2.27 | 1.79 | 1.93 | 1.74 | 2.11 | 1.97 | 1.88 | 2.06 | 2.14 | 1.92 |

Highest values and corresponding adjective in bold. Lower values and corresponding adjective in italics.

**Table 12**
Descriptive statistics for overall evaluation and purchase decision (study 2).

| | | Chair A-2 | | | Chair B-2 | | | Chair C-2 | | | Chair D-2 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | NPR | AR | VR | NPR | AR | VR | NPR | AR | VR | NPR | AR | VR |
| | Mean | *2.81* | 3.09 | **3.22** | *2.16* | 2.25 | **2.31** | 3.72 | *3.56* | **3.75** | *2.37* | **2.59** | 2.41 |
| Like/Dislike | Median | 3.00 | 3.00 | 3.50 | 2.00 | 2.00 | 2.00 | 4.00 | 4.00 | 4.00 | 2.00 | 2,50 | 2.00 |
| | SD | 0.96 | 1.09 | 1.13 | 0.99 | 1.05 | 1.15 | 0.96 | 1.10 | 1.05 | 1.21 | 1.36 | 1.32 |
| | Mean | *0.22* | 0.34 | **0.53** | *0.13* | **0.22** | 0.19 | **0.72** | *0.66* | **0.72** | *0.19* | **0.25** | *0.19* |
| Purchase decision | Median | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 |
| | SD | 0.42 | 0.48 | 0.51 | 0.34 | 0.42 | 0.40 | 0.46 | 0.48 | 0.46 | 0.40 | 0.44 | 0.40 |

Highest values in bold. Lower values in italics.

**Table 13**
Descriptive statistics for response confidence (study 2).

| | | Chair A-2 | | | Chair B-2 | | | Chair C-2 | | | Chair D-2 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | NPR | AR | VR | NPR | AR | VR | NPR | AR | VR | NPR | AR | VR |
| | Mean | *3.47* | 3.87 | **4.09** | *3.50* | 3.81 | **3.94** | 4.09 | *3.94* | **4,13** | *3.72* | 3.94 | **4.06** |
| Response confidence | Median | 4.00 | 4.00 | 4.00 | 3.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 |
| | SD | 0.80 | 0.66 | 0.59 | 0.84 | 0.82 | 0.91 | 0.86 | 0.80 | 0.79 | 0.92 | 0.88 | 0.88 |

Highest values in bold. Lower values in italics.

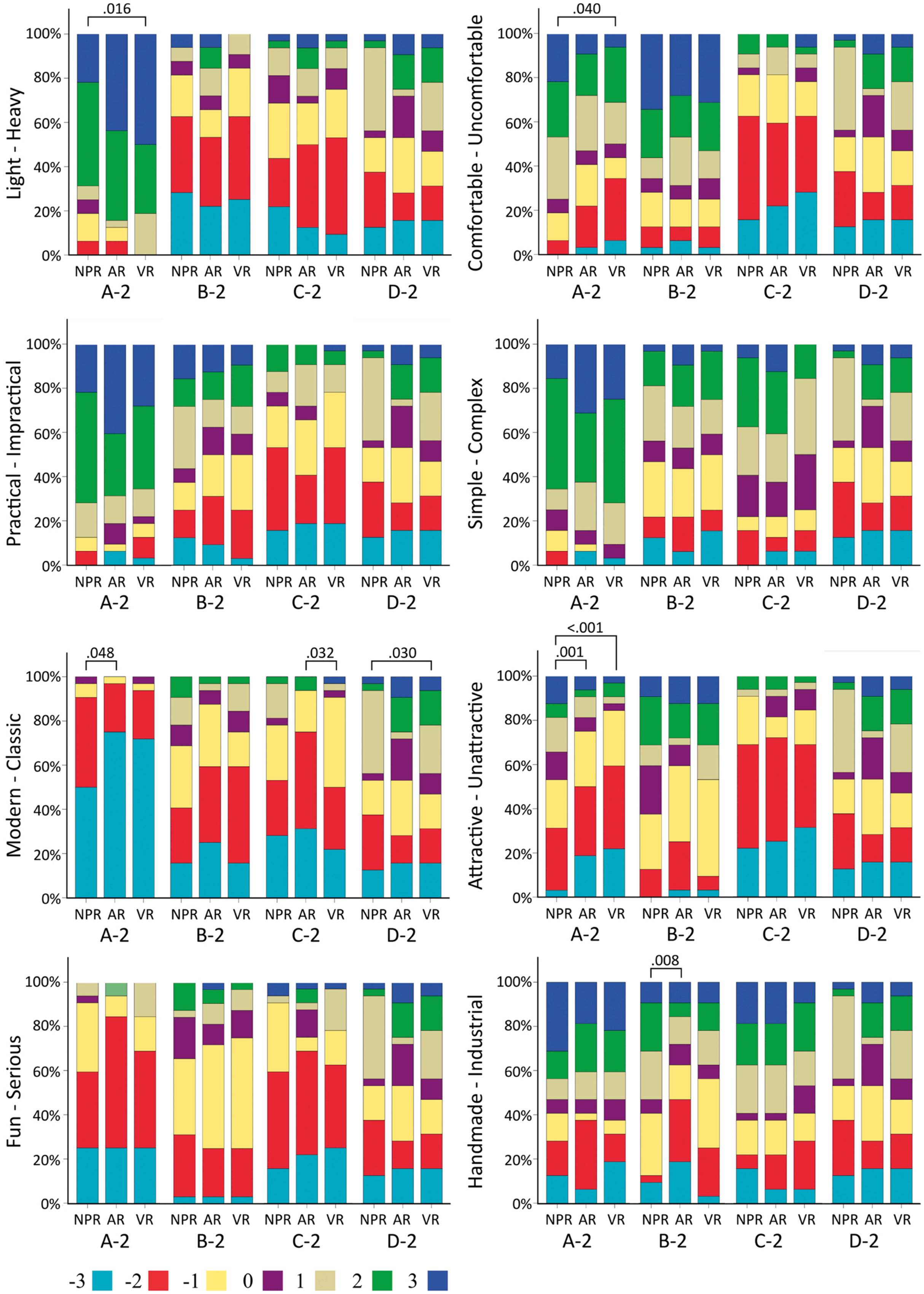


**Fig. 8.** Stacked bar charts for the semantic differential scales (study 2).

**Table 14**
Repeated Measures ANOVA for the semantic scales (study 2).

| | Semantic scales | df | Chair A-2 | | Chair B-2 | | Chair C-2 | | Chair D-2 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | F | Sig. | F | Sig. | F | Sig. | F | Sig. |
| PHYSIO | Light – Heavy | 2 | 5.024 | **p = .010** | 1.855 | p = .165 | .162 | p = .850 | .336 | p = .716 |
| | Comfortable – Uncomfortable | | 3.668 | **p = .031** | .153 | p = .859 | .202 | P = .817 | 1.446 | p = .243 |
| PSYCHO | Practical – Impractical | | .709 | p = .496 | .724 | p = .489 | .656 | p = .523 | 2.134 | p = .127 |
| | Simple – Complex | | 1.469 | p = .238 | .476 | p = .624 | 3.117 | p = .051 | 2.726 | p = .073 |
| SOCIO | Modern – Classic | | 3.445 | **p = .038** | 3.235 | **p = .046** | 3.697 | **p = .030** | 4.884 | **p = .011** |
| | Attractive – Unattractive | | 12.830 | **p < .001** | 1.519 | p = .227 | .205 | p = .815 | 1.236 | p = .297 |
| IDEO | Fun – Serious | | 1.239 | p = .297 | .072 | p = .930 | .146 | p = .865 | 1.560 | p = .218 |
| | Handmade – Industrial | | .577 | p = .564 | 5.492 | **p = .006** | .893 | p = 415 | .477 | p = .623 |

p-values less than.05 are shown in bold.

**Table 15**
Post–hoc tests for the semantic scales (study 2).

| Semantic scales | | Chair A-2 | Chair B-2 | Chair C-2 | Chair D-2 |
|---|---|---|---|---|---|
| | Means | Sig. | Sig. | Sig. | Sig. |
| Light - Heavy | NPR – AR | p = .072 | | | |
| | NPR - VR | **p = .016** | | | |
| | AR - VR | p = 1.000 | | | |
| Comfortable - Uncomfortable | NPR – AR | p = .083 | | | |
| | NPR - VR | **p = .040** | | | |
| | AR - VR | p = 1.000 | | | |
| Modern – Classic | NPR – AR | **p = .048** | p = .080 | p = .137 | p = .129 |
| | NPR - VR | p = .197 | p = .267 | p = 1.000 | **p = .030** |
| | AR - VR | p = 1.000 | p = .972 | **p = .032** | p = .530 |
| Attractive - Unattractive | NPR – AR | **p = .001** | | | |
| | NPR - VR | **p < .001** | | | |
| | AR - VR | p = .716 | | | |
| Handmade - Industrial | NPR – AR | | **p = .008** | | |
| | NPR - VR | | p = .554 | | |
| | AR - VR | | p = .155 | | |

p-values less than.05 are shown in bold.

anxiety: an individual's concern about being able to use a technological device correctly (Arvanitis et al., 2009). Users who have this anxiety are not likely to experience the AR in mobile applications effectively (Oyman et al., 2022), which may have influenced our results. The overall evaluation data set also showed statistically significant differences between means. These differences where found between IMG-AR and 3D-AR, expected result as AR was the only medium showing possible usage environment (the real world). Our results contradict those obtained by (Galán et al., 2021b), who compared the overall evaluation of one chair in a real setting, VR and VR with passive haptics. In our case, the visual media used to display the stimuli presented greater interaction differences, which may explain the results, which are similar to those obtained by (Palacios-Ibáñez et al., 2022).

**Table 16**
McNemar test for the purchasing decision (chair A-2).

| | Media | Chair A-2 Sig. |
|---|---|---|
| Purchase decision | NPR – AR | **p = .001** |
| | NPR – VR | **p < .001** |
| | AR - VR | p = .207 |

p-values less than.05 are shown in bold.

For the second study, results showed that participants were generally more confident in their responses in the VR environment. This result agrees with previous studies: while 2D media may offer sufficient information to evaluate a product, 3D representations provide richer information to the user (Ant Ozok and Komlodi, 2009), which helps them assess the appearance and features more directly (Liu, 2017), resulting in greater levels of certainty and confidence. Forbes et al. confirmed this result in their study with armchairs, where a more interactive and immersive medium helped to increase the user response confidence (Forbes et al., 2018). Although the overall evaluation data set did not show statistically significant differences between means, we observe that three of the four chairs were rated more favorably in VR, and all the designs were rated less favorably in NPR, as shown in the descriptive statistics for the overall evaluation. Therefore, we can confirm H0.3 and H0.4.

Finally, in H1.1, we speculated that a joint assessment could minimize the differences between the assessments made by the participants in different media. In the case of Chair C-1, the variation in the evaluation of the product is minimal after the change of media, according to (Galán et al., 2021b) (same score in both 3D and RA), as shown in Table 3 and Fig. 7. Likewise, variations in the purchasing decision responses are also minimal, as shown in Table 4. For example, for Chair A-1, the scores are the same for all three media, and for Chairs B-1 and

**Table 17**
Repeated measures ANOVA for the overall evaluation (study 2).

| | df | Chair A-2 | | Chair B-2 | | Chair C-2 | | Chair D-2 | |
|---|---|---|---|---|---|---|---|---|---|
| | | F | Sig. | F | Sig. | F | Sig. | F | Sig. |
| Overall evaluation | 2 | 2.780 | p = .070 | .300 | p = .734 | .800 | p = .455 | .860 | p = .428 |

**Table 18**
Cochran's Q test for the purchasing decision (study 2).

| | df | Chair A-2 | | Chair B-2 | | Chair C-2 | | Chair D-2 | |
|---|---|---|---|---|---|---|---|---|---|
| | | Q | Sig. | Q | Sig. | Q | Sig. | Q | Sig. |
| Purchase decision | 2 | 15.200 | **p < .001** | 2.800 | p = .247 | 0.800 | p = .670 | 1.330 | p = 513 |

p-values less than.05 are shown in bold.

**Table 19**
Post–hoc tests for response confidence (chair A-2).

| | Media | Chair A-2 Sig. | Chair D-2 Sig. |
|---|---|---|---|
| Response confidence | NPR – AR | **p = .020** | p = .298 |
| | NPR – VR | **p = .003** | **p = .031** |
| | AR - VR | p = .433 | p = .931 |

p-values less than.05 are shown in bold.

D-1, the same scores were obtained for two media (IMG – 3D and IMG – AR, respectively). Although the standard deviation in this dataset is not small and mean values may be misleading, they are smaller than those obtained in the second case study (Table 11) Our results thus confirm hypothesis H1.1.

We highlight that differences were found in Jordan's physio-pleasure category for both studies, and that the sociological pleasure category may also be affected by the change of medium if the products are highly aesthetical. It is interesting to note that for a simultaneous evaluation, users take less time to complete the task compared to an individual evaluation, and that the results obtained can be similar in both methods. Moreover, a simultaneous evaluation helps to minimize these differences. Therefore, we propose this method of evaluation as an effective alternative to evaluate aspects related to people's cognitive and emotional reactions (psycho and ideo pleasure categories), which was confirmed by other studies, such as Lee et al. (2004), who showed that the relative evaluation of the selected stimulus was the same regardless of the media used.

## 7. Conclusions

Being able to present a product effectively and understanding how it is perceived and assessed by users are critical factors for its success. Our study demonstrates that the visual medium used to present a product can influence how it is perceived and evaluated. Our results contribute to the literature of product design and engineering by empirically assessing the reliability of XR as a tool for product evaluation in the early stages of the product development process.

Certain characteristics such as comfort and size are particularly significant, as the perceptual differences elicited by different media are more pronounced. Although some studies highlight the importance of touch on the evaluation of a product, our results also show how other features that do not require haptics may also be affected by the influence of media or geometric product features (such as Jordan's socio-pleasure category). Our results have also revealed how these perceptual differences can be minimized, to a certain extent, by using joint product evaluation. In addition, certain product attributes can be emphasized in more immersive media (such as AR or VR), which is useful for both product development and point-of-sale presentation. We also emphasize the importance of having experience with AR and VR if the user will be using these technologies alone for product assessment, so that technology related anxiety does not negatively influence the evaluation.

We acknowledge the limitations of our study. First, all participants had limited experience in the use of AR, which may have influenced the evaluation of the products, especially in cases where the debriefing and follow-up interview was conducted online, instead of in person. Second, although our findings could potentially be extrapolated to similar products of the same typology, additional tests with other types of products are recommended to obtain more conclusive results. In future studies, we plan to conduct a similar experiment by changing the evaluation method (an individual evaluation for the first case study, and a simultaneous for the second). We also plan to use physiological measures such as eye-tracking technologies to analyze the user's gaze and overall behavior more accurately and objectively during product evaluation.

## Funding

This research work has been supported by the Spanish Ministry of Education and Vocational Training under a FPU fellowship (FPU19/03878). Funding for open access charge: Universitat Politècnica de València.

## CRediT authorship contribution statement

Conceptualization: **A. Palacios-Ibáñez, R. Navarro-Martínez, J. Blasco-Esteban, M. Contero**. Methodology: **A. Palacios-Ibáñez, R. Navarro-Martínez, J. Blasco-Esteban, M. Contero**. Software: **A. Palacios-Ibáñez.** Validation: **A. Palacios-Ibáñez, M. Contero**. Application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize study data: **A. Palacios-Ibáñez**. Investigation: **A. Palacios-Ibáñez, R. Navarro-Martínez, J. Blasco-Esteban, M. Contero, J. D. Camba**. Resources: **A. Palacios-Ibáñez, R. Navarro-Martínez, J. Blasco-Esteban, M. Contero**. Writing – original draft: **A. Palacios-Ibáñez**. Review & Editing: **M. Contero, J. D. Camba**. Visualization: **A. Palacios-Ibáñez**. Supervision: **M. Contero, J. D. Camba**. Project administration: **A. Palacios-Ibáñez, M. Contero**.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

The authors do not have permission to share data.

## Acknowledgements

The authors would like to thank the team at Clon Digital for providing us with a software license to perform the experiment, and students Jenny Trieu, Abizer Raja, Arturo Barrera, and Carrah Kaijser from the University of Houston for the inspiration for the chair designs used in our study.

## References

Aftab, M., Rusli, H.A., 2017. Designing visceral, behavioural and reflective products. Chin. J. Mech. Eng. (Engl. Ed.) 30 (5), 1058–1068. https://doi.org/10.1007/s10033-017-0161-x.

Achiche, S., Maier, A., Milanova, K., Vadean, A., 2014. Designing visceral, behavioural and reflective products. Chin. J. Mech. Eng. (Engl. Ed.) 11, 1–10. https://doi.org/10.1115/IMECE2014-40443.

Agost, M.-J., Vergara, M., Bayarri, V., 2021. The use of new presentation technologies in electronic sales environments and their influence on product perception. In: Proceedings of the International Conference on Human-Computer Interaction 12765, 3–15. https://doi.org/10.1007/978-3-030-78321-1.

Ant Ozok, A., Komlodi, A., 2009. Better in 3D? an empirical investigation of user satisfaction and preferences concerning two-dimensional and three-dimensional product representations in business-to-consumer e-commerce. Int. J. Hum. Comput. Interact. 25 (4), 243–281. https://doi.org/10.1080/10447310802546724.

Arbeláez, J.C., Osorio-Gómez, G., 2018. Crowdsourcing augmented reality environment (CARE) for aesthetic evaluation of products in conceptual stage. Comput. Ind. 99, 241–252. https://doi.org/10.1016/j.compind.2018.03.028.

Artacho-Ramírez, M.A., Diego-Mas, J.A., Alcaide-Marzal, J., 2008. Influence of the mode of graphical representation on the perception of product aesthetic and emotional features: an exploratory study. Int. J. Ind. Ergon. 38 (11–12), 942–952. https://doi.org/10.1016/j.ergon.2008.02.020.

Arvanitis, T.N., et al., 2009. Human factors and qualitative pedagogical evaluation of a mobile augmented reality system for science education used by learners with physical disabilities. Pers. Ubiquitous Comput. 13 (3), 243–250. https://doi.org/10.1007/s00779-007-0187-7.

Berni, A., Borgianni, Y., 2020. Applications of virtual reality in engineering and product design: why, what, how, when and where. Electronics 9 (7), 1–29. https://doi.org/10.3390/electronics9071064.

Berni, A., Maccioni, L., Borgianni, Y., 2020. An eye-tracking supported investigation into the role of forms of representation on design evaluations and affordances of original product features. Proc. Des. Soc. Des. Conf. 1, 1607–1616. https://doi.org/10.1017/dsd.2020.296.

Bordegoni, M., 2011. Product Virtualization: an Effective Method for the Evaluation of Concept Design of New Products. Springer, London.

Bruno, F., Muzzupappa, M., 2010. Product interface design: a participatory approach based on virtual reality. Int. J. Hum. Comput. Stud. 68 (5), 254–269. https://doi.org/10.1016/j.ijhcs.2009.12.004.

Cascini, G., et al., 2020. Exploring the use of AR technology for co-creative product and packaging design. Comput. Ind. 123, 103308 https://doi.org/10.1016/j.compind.2020.103308.

Cecil, J., Kanchanapiboon, A., 2007. Virtual engineering approaches in product and process design. Int. J. Adv. Manuf. Technol. 31 (9–10), 846–856. https://doi.org/10.1007/s00170-005-0267-7.

Christopoulos, G., Kokkinaki, F., Harvey, N., Sevdalis, N., 2011. Paying for no reason? (Mis-)perceptions of product attributes in separate vs. joint product evaluation. J. Econ. Psychol. 32 (5), 857–864. https://doi.org/10.1016/j.joep.2011.05.003.

Chu, C.H., Chen, Y.A., Huang, Y.Y., Lee, Y.J., 2022. A comparative study of virtual footwear try-on applications in virtual and augmented reality. J. Comput. Inf. Sci. Eng. 22 (4) https://doi.org/10.1115/1.4053328.

Chuang, M.C., Chang, C.C., Hsu, S.H., 2001. Perceptual factors underlying user preferences toward product form of mobile phones. Int. J. Ind. Ergon. 27 (4), 247–258. https://doi.org/10.1016/S0169-8141(00)00054-8.

Cooper, R.G., 2019. The drivers of success in new-product development. Ind. Mark. Manag. 76, 36–47. https://doi.org/10.1016/j.indmarman.2018.07.005.

Coutts, E.R., Wodehouse, A., Robertson, J., 2019. A comparison of contemporary prototyping methods. Proc. Int. Conf. Eng. Des. ICED 1313–1321. https://doi.org/10.1017/dsi.2019.137.

De Crescenzio, F., Bagassi, S., Asfaux, S., Lawson, N., 2019. Human centred design and evaluation of cabin interiors for business jet aircraft in virtual reality. Int. J. Interact. Des. Manuf. 13 (2), 761–772. https://doi.org/10.1007/s12008-019-00565-8.

Desmet, P., 2002. Designing emotions, reviewed by Norman. Des. J. 6 (2), 2–4.

Desmet, P.M.A., Ortíz Nicolás, J.C., Schoormans, J.P., 2008. Product personality in physical interaction. Des. Stud. 29 (5), 458–477. https://doi.org/10.1016/j.destud.2008.06.003.

F. Eggers, H. Sattler, T. Teichert, Choice-Based Conjoint Analysis. Springer International Publishing, 2022.

Faul, F., Erdfelder, E., Lang, A.G., Buchner, A., 2007. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav. Res. Methods 39 (2), 175–191. https://doi.org/10.3758/BF03193146.

Felip, F., Galán, J., García-García, C., Mulet, E., 2019. Influence of presentation means on industrial product evaluations with potential users: a first study by comparing tangible virtual reality and presenting a product in a real setting. Virtual Real. 24 (3), 439–451. https://doi.org/10.1007/s10055-019-00406-9.

Flynn T. N., Marley, A. A. J., 1992. Worst Scaling : Theory and Methods,178–201.

Forbes, T., Barnes, H., Kinnell, P., Goh, M., 2018. A Study into the influence of visual prototyping methods and immersive technologies on the perception of abstract product properties. In: DS 91: Proceedings of NordDesign 2018, Linköping, Sweden, 2018.

Galán, J., Felip, F., García-García, C., Contero, M., 2021. The influence of haptics when assessing household products presented in different means: a comparative study in real setting, flat display, and virtual reality environments with and without passive haptics. J. Comput. Des. Eng. 8 (1), 330–342. https://doi.org/10.1093/jcde/qwaa081.

Galán, J., García-García, C., Felip, F., Contero, M., 2021. Does a presentation media influence the evaluation of consumer products? A comparative study to evaluate virtual reality, virtual reality with passive haptics and a real setting. Int. J. Interact. Multimed. Artif. Intell. 6 (6), 196. https://doi.org/10.9781/ijimai.2021.01.001.

Hannah, R., Joshi, S., Summers, J.D., 2012. A user study of interpretability of engineering design representations. J. Eng. Des. 23 (6), 443–468. https://doi.org/10.1080/09544828.2011.615302.

Higgins, J.J., Blair, R.C., Tashtoush, S., 1990. The aligned rank transform procedure. Conf. Appl. Stat. Agric. https://doi.org/10.4148/2475-7772.1443.

Hsee, C.K., Loewenstein, G.F., Blount, S., Bazerman, M.H., 1999. Preference reversals between joint and separate evaluations of options: a review and theoretical analysis. Psychol. Bull. 125 (5), 576–590. https://doi.org/10.1037/0033-2909.125.5.576.

Jerald, J., 2015. The VR Book. In: The VR Book. Association for Computing Machinery.

Jordan, P.W., 2002. Designing Pleasurable Products: an Introduction to the New Human Factors. CRC Press.

Kato, T., 2019. Verification of perception difference between actual space and VR space in car design. Int. J. Interact. Des. Manuf. 13 (3), 1233–1244. https://doi.org/10.1007/s12008-019-00568-5.

Lee, S., Chen, T., Kim, J., Kim, G.J., Han, Sungho, Pan, Z., 2004. Affective property evaluation of virtual product designs. IEEE Virtual Real. 2004, 207–292. https://doi.org/10.1109/VR.2004.1310076.

Liu, S., 2017. The effect of on-line product presentation: a comparative study between 3d and general product presentation. J. Ind. Distrib. Bus. 8 (5), 15–21. https://doi.org/10.13106/ijidb.2017.vol8.no5.15.

Mansouri, H., Paige, R.L., Surles, J.G., 2004. Aligned rank transform techniques for analysis of variance and multiple comparisons. Commun. Stat. Theory Methods 33 (9), 2217–2232. https://doi.org/10.1081/STA-200026599.

Marquis, J., Deeb, R.S., 2018. Roadmap to a successful product development. IEEE Eng. Manag. Rev. 46 (4), 51–58. https://doi.org/10.1109/EMR.2018.2884275.

O'Keefe, R.M., McEachern, T., 1998. Web-based customer decision support systems. Commun. ACM 41 (3), 71–78. https://doi.org/10.1145/272287.272300.

Osgood, C.E., Suci, G.J., Tannenbaum, P.H., 1957. The Measurement of the Meaning. University of Illinois Press, Champaign, IL.

Oyman, M., Bal, D., Ozer, S., 2022. Extending the technology acceptance model to explain how perceived augmented reality affects consumers' perceptions. Comput. Hum. Behav. 128, 107127 https://doi.org/10.1016/j.chb.2021.107127.

Palacios-Ibáñez, A., Ochando-Martí, F., Camba, J.D., Contero, M., 2022. The influence of the visualization modality on consumer perception: a case study on household products 49, 160–166. https://doi.org/10.54941/ahfe1002050.

Pleyers, G., Poncin, I., 2020. Non-immersive virtual reality technologies in real estate: how customer experience drives attitudes toward properties and the service provider. J. Retail. Consum. Serv. 57, 102175 https://doi.org/10.1016/j.jretconser.2020.102175.

Söderman, M., 2005. Virtual reality in product evaluations with potential customers: an exploratory study comparing virtual reality with conventional product representations. J. Eng. Des. 16 (3), 311–328. https://doi.org/10.1080/09544820500128967.

Song, H., Chen, F., Peng, Q., Zhang, J., Gu, P., 2018. Improvement of user experience using virtual reality in open-architecture product design. Proc. Inst. Mech. Eng. Part B J. Eng. Manuf. 232 (13), 2264–2275. https://doi.org/10.1177/0954405417711736.

Ray, S., Choi, Y.M., 2017. Employing design representations for userfeedback in the product design lifecycle. In: Proceedings of the Twenty First Int. Conf. Eng. Des. ICED 4 (DS87-4), 563–572.

Statista, Augmented Reality (AR) and Virtual Reality (VR) Market Size Worldwide From 2016 to 2024, 2021. ⟨https://www.statista.com/statistics/591181/global-augmented-virtual-reality-market-size/⟩. (Accessed 2 March 2022).

Tran, L.T.T., 2021. Managing the effectiveness of e-commerce platforms in a pandemic. J. Retail. Consum. Serv. 58, 102287 https://doi.org/10.1016/j.jretconser.2020.102287.

Van Kleef, E., Van Trijp, H.C.M., Luning, P., 2005. Consumer research in the early stages of new product development: a critical review of methods and techniques. Food Qual. Prefer. 16 (3), 181–201. https://doi.org/10.1016/j.foodqual.2004.05.012.

Wang, Y., Jia, F., Schoenherr, T., Gong, Y., Chen, L., 2020. Cross-border e-commerce firms as supply chain integrators: the management of three flows. Ind. Mark. Manag. 89, 72–88. https://doi.org/10.1016/j.indmarman.2019.09.004.