



The Aesthetic Pleasure in Design Scale for Spanish Speaking Countries: A Method for the Cross-Cultural Implementation and Adaptation of Psychometric Scales

Luis Miguel Garrido-Possauner^(✉) and Jorge Maya

Universidad EAFIT, Medellin, Colombia
{lgarrido, jmayacas}@eafit.edu.co

Abstract. One of the main issues in the field of product aesthetics is the lack of consistency in the instruments and terms being used to assess it. Several scales have been used but with restricted validity and reliability; notwithstanding, the APID (Aesthetic Pleasure In Design) scale developed in English, as part of the UMA (Unified Model of Aesthetics) project, lacks these limitations. Our research, being in a Spanish speaking country, required the scale to be in Spanish to be comprised and applied to domestic respondents. The adaptation of psychometric instruments with conceptual and linguistic challenging constructs, often is difficult, as the items used to assess the construct do not always have a direct translation. Also, it is possible for the adapted version to require different items that were not considered in the original instrument in order to preserve the content's equivalence according to the target culture. A qualitative-quantitative mixed approach is proposed in order to overcome these issues. The resulting instrument has been statistically tested, proving to be both valid and reliable for the measurement of aesthetic pleasure in design. The final scale consists of five items: *bonito*, *hermoso*, *agradable*, *llamativo* and *me gusta*. This paper aims to improve the understanding of how aesthetic pleasure is perceived and therefore expressed by the local respondents gaining insight into how this construct is mentally represented and categorized by the respondents. It also aims to illustrate how psychometric scales based on respondents' vocabulary have a great potential as usability assessment instruments.

Keywords: Cross-cultural research instruments · Cultural differences · Scale adaptation

1 Introduction

1.1 Aesthetics' Assessment

The study of aesthetics has been a topic of growing relevance in the past years. Consequently, there is an increasing interest in trying to improve the current understanding of aesthetics in design and the factors that relate to it. Most studies in the field of product design are focused on the determinants, rather than in the study of aesthetic pleasure itself, because determinants are the variables that can be directly modified by

designers (Blijlevens et al. 2017). It would be nonsense to deny the apparent universality of some aesthetic preferences as the gestalt principles (Hekkert and Leder 2008; Shortess et al. 1997; Fechner 1997), but it could also be harmful to assume them as the only influencing factor. Although many studies have successfully tested the determinants' influence on people's perception of products (Roussos and Dentsoras 2013; Hekkert and Leder 2008), judging a product's aesthetics by evaluating only the determinants has its limitations, as their ability to predict the aesthetic response to a stimulus, is time- (Jacobsen 2010) cultural- (Hekkert and Leder 2008) and product category- (Hekkert 2015) dependent. Not all the determinants have the same influence in every context (Berghman and Hekkert 2017; Hekkert 2015). As Oscar Wilde, said "no object is so beautiful that, under certain conditions, it will not look ugly". Most of the empirical studies regarding determinants have been tested on a specific product category, on a single context, and consequently, the methodological differences between studies make it difficult to compare studies across different contexts and product categories (Blijlevens et al. 2017).

Another way to evaluate product aesthetics is to measure the response from the target audience. This aesthetic response is considered to be intangible and therefore latent, as it cannot be directly observed (Blijlevens et al. 2017; Jöreskog and Sörbom 1979). Psychometric instruments have been widely used to assess latent constructs in social sciences, medical care and other fields. However, when it comes to using a psychometric instrument developed in one language, in a translated version, problems might arise as a result of a poor translation process. This paper aims to explain the process behind the adaptation and implementation of an aesthetic pleasure in design scale, originally in English, to a Spanish speaking country.

1.2 Aesthetic Pleasure

Defining the Construct. For many years, researchers have mentioned the existence of a unique underlying factor behind the aesthetic experience (Eysenck 1940; Marty et al. 2003). Despite this, research on how to define and measure the aesthetic pleasure as a construct of interest has received little attention (Blijlevens et al. 2017). For our research, aesthetic pleasure will be understood as the "sensorial **pleasure and delight**" (Goldman, 1990) "people derive **from processing** the object for its own sake, as a source of **immediate experiential pleasure in itself**, and not essentially for its utility in producing something else that is either useful or pleasurable" (Dutton 2009, p. 52).

Measuring Aesthetic Pleasure. Many scales have been used to measure aesthetic appreciation (Faerber et al. 2010; Page and Herr 2002; Hung and Chen 2012; Martindale et al. 1990; Hassenzahl and Monk 2010), but with a lack of reliability or validity (Blijlevens et al. 2017). The biggest concerns in this area are summarized in three aspects. One, determinants and/or semantic descriptors are often used inside the scales, which makes it hard to isolate the measurement of the aesthetic response and, therefore, noise is generated in the assessment (Blijlevens et al. 2017; Faerber et al. 2010). Two, the lack of consistency between studies, as the scales used differ from one research project to another, making it difficult to make comparisons between studies

(Blijlevens et al. 2017). This lack of precision in terminology is one of the biggest problems concerning literature on psychological aesthetics, as mentioned by Augustin et al. (Augustin et al. 2012; Faerber et al. 2010). Three, the instruments used in these studies are often used ad hoc or without a mention of their origin or validity (Blijlevens et al. 2017).

In the Latin American context, the problem is even bigger as the number of instruments developed in Spanish is much lower. Many studies use translated items without a former validation. Spanish is an official language in more than 20 countries for more than 400 millions of people (Stewart 2003) and the difference in the lexicon and usage of the language among different countries and cultures makes it difficult to establish a “standardized language”. There is no evidence of a specific instrument for measuring aesthetic pleasure in product design in Spanish. Hernández Belver (1989) carried out a study in which he included the set of items *bonito – feo* (beautiful – ugly), *agradable – desagradable* (pleasing – unpleasing) and *interesante – no-interesante* (interesting – not interesting) to rate artistic related stimuli. Later, Marty et al. (2003) implemented Belver’s items after adding the new pair of items *original – común* (original – common), which are actually known determinants (Hekkert and Leder 2008; Berghman and Hekkert 2017), rather than aesthetic responses. Both studies propose interesting items but lack a strong theoretical background. Marty et al. performed a factor analysis inside their study, but its scope was completely exploratory as their objective was to explore search for empirical evidence of an underlying factor behind the aesthetic experience (Marty et al. 2003). Also, the item generation was based on only a few authors’ work.

The APID (Aesthetic Pleasure In Design) scale was developed in English as part of the UMA project (Unified Model of Aesthetics). This project aims, as its name suggests, to unify the different theories behind the explanation of aesthetic pleasure (Berghman and Hekkert 2017). This project was developed inside a design-oriented research. Stimuli from different product categories were used to improve the instrument’s robustness. The instrument consists of five items (beautiful, attractive, pleasing to see, like to look and nice to look). This scale has been tested proving to be a valid and reliable instrument (Blijlevens et al. 2017). Because of its psychometric properties and its strong theoretical background, it was identified as an ideal instrument to implement in our research. But, in order to be able to use the APID scale in the local context, the instrument had to be in the local language first.

1.3 Translating the APID Scale

Translation of Psychometric Instruments. Translation is the act of rendering knowledge available from one culture to another (Montgomery 2006). Implementing a scale for its use in a different culture is a process that often requires considerable effort by researchers (Brislin 1970; Wang et al. 2006). Contrary to the translation of a text, translating a measurement scale does have rules for correctness (Montgomery 2006). These rules are the same used in the construction of the scale in the source language (see Blijlevens et al. 2017). In other words, it is necessary to follow, in the translation, a method to ensure that the scale in the target language fits all the requirements a scale

must fit. Equivalence between the original and the translated version of the instrument must be preserved. In translation studies, equivalence is the notion used to explain what in natural sciences is called precision. The notion of equivalence means that the objective of a translation “is to produce a target language text which is equivalent to the original language text.” p. 86 (Sequeiros 2006). According to the Webster dictionary, equivalent means “having the same or similar effect or meaning”. In the past decade, there have been increasing numbers of publications on translating and adapting instruments from one culture to another. Eremenco et al. (2005) identified 5 types of equivalence:

1. Content Equivalence: each item’s content is relevant in both cultures;
2. Semantic Equivalence: the similarity of meaning of the items in both cultures after translation is emphasized;
3. Technical Equivalence: data collection methods for the 2 versions of the instrument are similar;
4. Criterion Equivalence: scores are interpreted in the same way in their respective cultures;
5. Conceptual Equivalence: the instrument measures the same theoretical construct in each culture.

The so called “free” and “literal” approaches to translation cannot be used in our case because the researchers cannot see if the translated terms really fit (Montgomery 2006). Moreover, “one-to-one correspondence in scientific translation does not exist.” (Montgomery 2006, p. 67). Many methods have been proposed in order to protect the equivalence. Here, in Table 1, are some of the most commonly used methods for translating instruments (Brislin 1970).

Table 1. Typically used instrument translation methods

Method	Description	Advantages	Disadvantages
Back-translation	First, a bilingual translator translates an instrument from the original language into the target language; Second, another independent bilingual translator translates the instrument from the target language back into the original language. The two versions of the tool are then compared for concept equivalence. When an error is found in the back-translation, another translator attempts to re-translate it. This procedure continues until the team of bilingual translators agrees that both versions (the	Back translation provides a way to systematically contrast the translated version to the original items.	Some items are more linguistically and conceptually challenging and it might be reflected in a higher difficulty for the experts (Hilton and Skrutkowski 2002) Bilingual translators may be influenced by the source culture when doing the translation (Hilton and Skrutkowski 2002) Translators’ lack of expertise in the instrument’s knowledge area may affect the equivalence of the translated version

(continued)

Table 1. (continued)

Method	Description	Advantages	Disadvantages
	original and its translation) are equivalent (Cha et al. 2007; Brislin 1970; Behling and Law 2000)		Although back-translation is used, the translated version may not be appropriate for its use with the target population, especially for cross-cultural research. (Maneesriwongul and Dixon 2004)
Bilingual technique	Two versions of the same instrument, one in the original and one in the target language, are administered to a group of bilingual participants. Responses to both version are then compared to identify discrepancies (Brislin 1970)	Allows immediate respondent's feedback	Requires a group of bilingual respondents, which is hard to obtain in the local context Bilingual respondents might have a bias regarding their acculturation and may therefore answer differently than local respondents would. (McDermott and Palchanes 1992; Sperber et al. 1994) This method completely relies on the translators' interpretation of the reasons behind the discrepancies
Committee approach	This approach is based on the use of groups as a way to mitigate possible personal mistakes in the translation process. It consists of the implementation of a committee, in which bilingual experts translate the instrument as a team. (Brislin 1970)	The collaborative approach, helps to eliminate time delays between iterations It helps to prevent possible bias from individuals	This method requires a minimum of three bilingual people. Accessibility of bilingual people as translators is a key issue when applying this approach
Pretest procedure	It consists of the implementation of a pilot test, which allows researchers to identify potential issues with the level of understanding of the translated instrument, by testing it with a smaller group (Brislin 1970)	It allows to have a quick look at people's level of understanding of the instrument	Pre-test respondents might not be representative of the target context This method is used to give an insight on possible issues but it is not a translation method by itself

No matter which method is being used, interpretation is always present in translation (Montgomery 2006); the translation of a scale is a question of equivalence though; consequently, interpretation should be ruled out as far as it introduces the translator's subjectivity.

Translating the Scale. The first step in our research was the translation of the reference instrument. As the study's resources and availability of qualified translators

were limited, the research team decided to implement a pretest method as a first approach, as it allows researchers to have a diagnosis of possible misunderstandings and mistakes, which is important given that the items have a high loading of cultural content. The original items were translated by two certified translators who used an intra-translation process, meaning that both worked separately and did not see any of the other translator's work until the end of the process, this in order to maximize the variability. Both translators have Spanish as their main language, but one of them is Colombian and the other one Spanish.

Table 2. Technically translated versions of the APID scale

Original item	Translator 1	Translator 2
This is a beautiful product	Este es un excelente/hermoso/bonito producto	Este es un producto bonito
This is an attractive product	Este es un producto atractivo	Este es un producto atractivo
This product is pleasing to see	Este producto es agradable a la vista	Ver este producto es un placer
This product is nice to see	Da gusto ver este producto	Este producto es visualmente agradable
I like to look at this product	Me gusta ver este producto	Me gusta mirar este producto

As seen in Table 2 both translators reached a similar outcome with small differences. Consequently, the items chosen for the initial translated version were *bonito*, *atractivo*, *agradable a la vista*, *da gusto ver* and *me gusta ver*. The items were then pre-tested with a small group of people, ten students from the university. This pre-test procedure was not performed systematically and the information gathered was completely qualitative rather than quantitative. Participants were asked to give comments while rating a small number of products with the translated version of the scale. Even though no large discrepancies between the translators' versions had been found, the pretest of the translated version showed that respondents did not understand the instrument and its items properly. It has been noted that data obtained from the general population "are best when the question is clear, and when the respondent knows the answer and is motivated to report it accurately" (Mechanic 1989, p. 150). As mentioned before, items should not only represent theoretical meaning, but they should also be a reflection of how people actually express themselves. For instance, "beautiful" and "attractive" could be defined as two wholly different words according to the dictionary; however, for Colombian people, using them as two different items while assessing a product was difficult. Besides, it is usual in English to define a product as "attractive", but is fairly uncommon in Colombian Spanish to use the words "*atractivo*" or "*atrayerente*" with that purpose. This was evidenced before by Deutcher (1973), as he

noted that even though a translation is considered equivalent in a back-translation process (e.g. *amigo*, *ami*, *tomodachi* for the word “friend”), the original and translated version of the same word may have important differences in their linguistic nuances. Roughly speaking, in a translation task, a semantic network (of the source language term) is activated in the source language; this activated network also includes nodes for the concept and highly salient structures in the target language which exert a “gravitational pull” resulting in an overrepresentation in the translated terms of those salient structures (Halverson 2003). Even if the conceptual and linguistic equivalence is assured, the measurement of the same concept may require different items or indicators across cultures (DeVos 1973). There is also growing evidence that the experiences, expressions, and correlates (e.g. for depressive disorders) are not universal but rather vary depending on the ethno cultural context (Marsella et al. 1973, 1987). The assumption of cultural universality in the construction of research instruments may lead to an inadequate implementation and even an erroneous interpretation of the research findings (González-Calvo et al. 1997).

Hines (1993) proposed a combination of qualitative and quantitative methods as a way to improve the quality of cross-cultural instruments as it helps researchers to create instruments that are more relevant to the target culture. He points out that the use of cognitive techniques may provide information to “better understand how different cultural and ethnic groups construe the world” (Hines 1993), as the information provided by these techniques correspond to the respondents’ underlying thought processes. Methods such as free listings, frames, rank orders, triad tests and pile sorts are recommended by the author. Other authors also state that the usage of such techniques may protect the instrument’s content validity in the target culture by being a source of relevant and appropriate items (González-Calvo et al. 1997). A qualitative-quantitative approach was then adopted by the research team, as this allowed us to overcome the previously found problems: (1) The lack of understanding of the instrument, due to the use of items that were irrelevant in the context; (2) The fact that different items might be needed in the translated version in order to protect the construct’s validity.

2 Methodology

Proposed Methodology. For our research, we propose a combination of Free Listing and Card Sorting as the fundamental activities in the item generation phase. The combination of these two methods has been previously proposed and used (González-Calvo et al. 1997; Sinha 2004), proving to be a successful way to explore and understand the respondents’ vocabulary and underlying mental models. After the item generation phase, an exploratory factor analysis will be performed in order to test the construct validity of the proposed adapted instrument (Table 3).

Table 3. Overview on the adaptation methodology

Phase	Method	Objective	Inputs	Output
1	Free listing	To identify local respondents' aesthetic relevant vocabulary, to create an initial set of items	Free Listing physical formats	Elicited items ranked by frequency of mention and cognitive saliency indexes
	Card sorting	To visualize the perceived connections between the elicited items in order to understand the relationships between terms and their perceived similarity.	Filtered elicited items	Clusters representing mental connections between items Selection of final items to validate
2	Exploratory factor analysis	To test the instrument validity and reliability.	Proposed items Respondents' evaluations of different products	Exploratory validation of the instrument Final set of items loading the construct of aesthetic pleasure

3 Procedure and Results

The methods for the initial item generation can be deductive or inductive. Deductive methods are based on an extensive literature review and the study of pre-existing scales (Hinkin 1995), while inductive methods are based on qualitative data gathered from the target population (Kapuscinski and Masters 2010). In this case, Free listing and Card Sorting will be used as inductive methods to generate and understand relevant items and their connections.

3.1 Step One: Free Listing

Free listing is an elicitation technique that has been widely used inside the social sciences (Hines 1993). An example of its usage in scale development processes can be evidenced in Kinzie et al. study, as it was used to gather information in the item generation phase for the development of the Vietnamese-language Depression Rating Scale (Kinzie et al. 1982). Free Listing allows researchers to get a better understanding of the knowledge a group of people has about a particular subject and the vocabulary they use to make reference to it. Free Listing is a simple but structured method, which allows researchers to have access to a lot of information about the cultural domain. A cultural domain, as defined by Borgatti (1999), is a set of concepts that seem to belong to the same mental group or category for a specific cultural group. This method consists of asking participants to “list all the adjectives and words of X that they can think of”. According to Smith (1999), items with a higher frequency and average position within the lists are the most relevant ones for the target group.

Respondents. A total amount of 332 participants took part in this study, which was conducted in Medellín, Colombia, and its surroundings. Respondents were selected by convenience with pre-defined quotas, as done before by Antmann et al. (2011). Four different companies (a domestic appliances manufacturer, a clothing manufacturer, a textile manufacturer and a bank service provider) collaborated in the study by allowing the research team to conduct the research activities with their employees. Employees from different areas, with different backgrounds, participated within each company. Students from the university also participated in the study. This combined strategy allowed us to reach a high level of diversity, while maintaining pre-defined criteria: One, all participant had Spanish as their main language. Two, none of the respondents had a design or art related job or profession, as the objective was to explore and understand the non-designers' and non-artists' aesthetic vocabulary. Three, the proportion of female and male respondents had to be similar. Of these 332 participants, answers were not considered from people who did not provide all the information (age, gender, main language). The final analysis was performed with a total of 270 participants (mean age 34, SD 14, 140 females).

Procedure. Respondents were asked to write down all the positive terms or expressions they could think of when asked to describe a product's aesthetics. A total amount of 342 items were collected. Lists were then analyzed with the ANTHROPAC software. Colloquialisms were not considered for further analysis in order to protect the scale's generalizability.

Table 4. Twenty elicited words with the highest scores regarding product aesthetics.

Item	Frequency (%)	Salience
Bonito (beautiful)	44.1	356
Lindo (cute/pretty)	20.4	137
Hermoso (gorgeous)	19.3	136
Bueno (good)	15.6	117
Elegante (elegant)	17.8	111
Agradable (nice/pleasant)	11.5	70
Excelente (excellent)	10.7	70
Espectacular (spectacular)	10.4	67
Calidad (quality)	9.6	65
Útil (useful)	10.0	59
Color (color)	9.3	57
Me gusta (I like it)	9.6	52
Práctico (practical)	8.5	52
Cómodo (comfortable)	7.8	50
Me encanta (I love it)	9.3	47
Belleza (beauty)	6.3	44
Rico (tasty)	7.4	42
Divino (divine/adorable)	7.8	41
Súper (super)	7.0	41
Llamativo (eye-catching)	5.9	41

*(Approximate Translations to English)

Most of the elicited items were directly related to aesthetic pleasure. However, as seen in other studies (Antmann et al. 2011), some of the words had more of a descriptive nature than an evaluative one, making reference to specific attributes such as color and size. Also, other words had a semantic nature (comfortable, modern, practical, economic) making reference to products’ meanings rather than the perceived aesthetic pleasure. As expected, *bonito* (beautiful) is by far the most relevant term in the domain, as seen before in other studies (Jacobsen et al. 2004; Augustin et al. 2012). However, the assumption of a single explanatory dimension (beautiful-ugly) could harm the construct’s content validity, given the huge amount of expressions used to describe it, as evidenced in the item list. Also, as seen in Table 4, the item attractive was not one of the main twenty elicited items, even more, it was one of the least mentioned items. The absence of this term confirms its low relevance when measuring a product’s aesthetics in the local context.

A threshold was established and the only items retained were those whose index was higher than the average. The elicited items were then filtered by the research team according to their coherence with the target construct according to the referenced theoretical background. Physical attributes (such as “colors”), semantic concepts (such as “modern”) and known determinants (such as “symmetric”) were removed from the list. Twelve items were selected as possible candidates for the proposed scale. Finally, six researchers with previous experience in the field of product aesthetics rated the remaining 12 items on the level to which they thought these items were representative of the construct aesthetic pleasure by using a web-based questionnaire. Coherence (the extent to which the term is directly related to the construct of aesthetic pleasure), practicality (the term is easy to understand and use) and relevance (the term’s appropriateness for its use within the field of product design) were establish as the assessing criteria in order to identify the item’s representativeness as done before by our reference scale (Blijlevens et al. 2017). All the experts were asked to rate the different items according to these criteria, using a five-point Likert scale. Items with a score higher than three were rounded up while items with a score lower than three were rounded down. Results are presented in Table 4.

Only the items with a score of three points or higher were retained for further analysis (Table 5).

Table 5. 1-least appropriate to measure aesthetic pleasure, 5-most appropriate to measure aesthetic pleasure.

1	2	3	4	5
Bueno	Elegante	Espectacular	Me gusta	Bonito
Excelente	Interesante	Divino	Me encanta	Lindo
Cómodo	Maravilloso		Llamativo	Hermoso
Rico	Delicioso		Estético	Agradable
Genial				Belleza
				Bello

3.2 Step Two: Card Sorting

Card sorting is a clustering method that allows researchers to identify respondents' levels of meaning and mental connections between concepts (Capra 2005; Hines 1993). Similar approaches have used Card Sorting to gain understanding of constructs such as automotive seat comfort (Erol 2018). "According to cognitive anthropologists, uncovering ways in which various cultural groups classify and divide concepts provides valuable insight into the way a particular group defines and organizes reality" (Hines 1993).

Respondents. A total of 24 respondents from Medellin, Colombia participated in this study. 1. All respondents were undergraduate students or workers. 2. None of the respondents has an art or design related job/profession. 3. The proportion of female and male participants had to be similar.

Procedure. Open Card sorting was selected as the best option as it allowed respondents to create their own groups without being biased by a pre-established structure (Spencer and Garrett 2009) Participants were asked to group the different terms according to their similarity. The resulting 12 items from the previous method were used as input. The number of groups was not limited to a maximum or minimum. Data was analyzed using the SynCaps software.

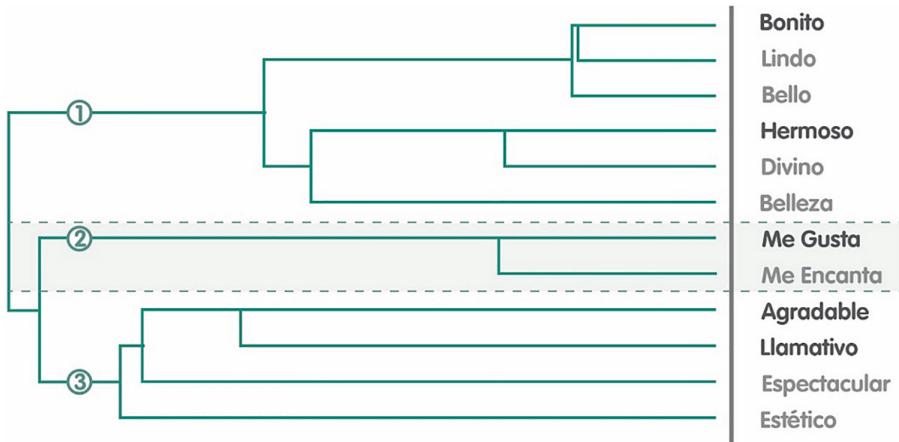


Fig. 1. Cluster analysis after card sorting.

After the cluster analysis three main groups were identified. The research team decided that at least one item from each cluster should be selected in order to have a good content validity. Bonito, me gusta and agradable were then selected as the representative items for each cluster, as they had the highest frequency of mention and its usage was the least age and gender dependent. Hermoso was also selected for further analysis as the research team thought it could provide different information than the

item bonito. Many aesthetic theories show the importance of both, the presence of interest and an aesthetic liking inside the aesthetic experience (Graf and Landwehr 2017; Berlyne 1971). Martin also states, from a completely different background, that inside appraisals systems in language there are two different types of reaction when appreciating a stimulus (Martin 2000): impact and quality. Impact is related to the capacity a stimulus has when it comes to captivating the preceptor's attention, while quality describes the positive effect it transmits. For this reason, *llamativo* (eye-catching) was also selected to be part of the proposed scale, as the research team believed the scale's content validity could be harmed if left out.

As a result, five items are proposed to assess aesthetic pleasure in the local context: *bonito* (beautiful), *hermoso* (gorgeous), *agradable* (nice/pleasing), *llamativo* (eye-catching) and *me gusta* (I like it).

4 Phase Two: Exploratory Factor Analysis

EFA is a commonly used statistical method that allows researchers to evaluate the construct validity of a scale, test, or instrument (Pett et al. 2003; Thompson 2004) by verifying if the proposed items are actually driven by the same underlying latent variables (Field 2009).

4.1 Method

Stimuli Selection. Twenty product images were selected to be rated by the participants. As done by Blijlevens et al. (2017), four different product categories were chosen as stimuli (cameras, motorcycles, chairs, and websites) to improve the generalizability of the scale across different product categories. Five different designs were selected to represent the variety found within each product category. All the stimuli were presented in the same layout, preserving perceptual equivalence. No renders nor concepts were used but rather a photography of the real product. Identifying brand features were removed from the images in order to avoid possible brand related bias.

Respondents. Respondents were recruited by convenience from different contexts and backgrounds in order to keep the sample as heterogeneous as possible. All the respondent's answers that had only extreme scores (1 or 7), neutrals (4), or consecutive responses (e.g., 3, 3, 3...) were deleted before the analysis. Answers in which more than 50% of the scores were assigned to the same value and incomplete answers were also deleted. The final analyses were performed with a total of 142 respondents (85.9% between 19 and 45 years old, 95 females). A minimum amount of 10 respondents per item is recommended while more than 15 is considered ideal (Clark and Watson 1995; DeVellis 2003; Hair Junior et al. 2009).

Procedure. Respondents were asked to view and rate a series of images of products. They were asked to indicate the extent to which they agreed with different statements describing each given stimulus by using a 7-point Likert scale (1 strongly disagree, 7 strongly agree). The aforementioned final items from the generation phase were used

for aesthetic pleasure. Items were stated in a way they made a judgment over the stimuli rather than the action (For example, I like the way this product looks, rather than, I like to look at this product). Three items, *novedoso* (*novel*), *innovador* (*innovative*) and *original* (*original*), representing the determinant novelty, and one item, *feo* (*ugly*), representing a commonly used opposite term, were used to assess the discriminant validity of the aesthetic pleasure scale. The questionnaire was created using the web platform Typeforms. This tool is time efficient, as it immediately shows the next question after the participant chooses an answer without having to scroll down. However, this platform does not allow a randomization in the question order so four versions of the questionnaire (with completely different orders) were created to reduce possible bias.

4.2 Results

Correlation Matrix. As the first step of the validation phase a correlation matrix between the nine items was created. Darker colors represent higher correlations between items (closer to 1 or -1). Blue stands for positive correlations while red stands for negatives.

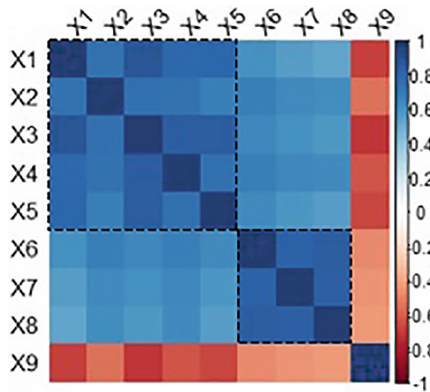


Fig. 2. Correlation matrix – variables measuring aesthetic pleasure and novelty. (Color figure online)

A correlation matrix serves as an indicator to researchers to identify variables that cluster together. “This data reduction is achieved by looking for variables that correlate highly with a group of other variables, but do not correlate with variables outside of that group” (Field et al. 2014). As evidenced in Fig. 1, there are two main data groups that can immediately be recognized. This is a great indicator as the first five variables are the ones supposed to measure aesthetic pleasure and variables X6 to X8 are the ones measuring the determinant “novelty”. All correlations within these two groups were above 0,68 and there was not a correlation above 0,9, which means that redundancy was not present. The fact that variables measuring novelty have a medium-high

positive correlation with the variables measuring aesthetic pleasure prove the selected items to be a good contrast for discriminant validity. Variable X9, which stands for *feo* (ugly) was also intended to allow a look at the discriminant validity, by standing for an opposite measure. This variable showed a negative correlation to all the variables. As expected, these correlations reach the highest scores when compared to the variables measuring aesthetic pleasure.

PCA and Cluster analysis. A principal component analysis (PCA) was conducted on the 9 items. In order to verify the sampling adequacy for factor analysis, a Kaiser–Meyer–Olkin measure, KMO, was performed. A KMO value of .934 was obtained, which is considered “superb (Field et al. 2014). All KMO values for individual items were $>.88$, which is highly above the acceptable limit of .5 (Field et al. 2014). Bartlett’s test of sphericity = 27102.23, $p < .001$, indicated that the correlations between items were sufficient for PCA (Bartlett 1950). Eigenvalues for each component in the data were obtained. Three components had eigenvalues over Kaiser’s criterion of 1 (Kaiser 1960), explaining, in combination, **87.7%** of the variance. The scree plot showed inflexions that would justify retaining both 2 or 3 components, but as Fig. 3 shows, it was clear that the third component was just an opposite manifestation of the first component as it is located on the same diagonal.

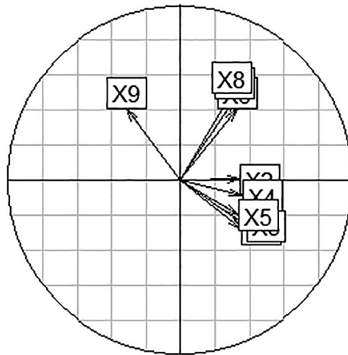


Fig. 3. Principal components PC1, PC2, PC3.

Exploratory Factor Analysis. An exploratory factor analysis with oblique rotation, promax, was performed in order to enhance the differentiation, as the variables loading the different factors were known to have a correlation. Table 2 shows the factor loadings after rotation. The items that cluster on the same factor suggest that factor 1 represents aesthetic pleasure and factor 2 the determinant of aesthetic pleasure, novelty (Table 6).

Table 6. Variable loadings after EFA with oblique rotation

Loadings	Factor 1	Factor 2
<i>Este producto es bonito</i> (this is a beautiful product)	0.96	
<i>Este producto es llamativo</i> (this product is eye-catching)	0.60	
Me gusta como se ve este producto (I like how this product looks)	0.98	
<i>Este producto es hermoso</i> (this is a gorgeous product)	0.74	
Este producto es agradable (this is a nice product)	0.84	
<i>Este producto es feo</i> (this is an ugly product)	-0.79	
Este producto es novedoso (this is a novel product)		0.85
<i>Este producto es original</i> (this is an original product)		0.89
<i>Este producto es innovador</i> (this is a innovative product)		0.99

The five items were retained after the exploratory factor analysis.

Reliability. Cluster analysis revealed that all correlations were above .50 and significant, so all items were retained. Factor in-variance analysis showed no evidence of significant discrepancies between product categories for each factor. Cronbach's alphas were .70 for aesthetic pleasure and .83 for novelty. To assess retest reliability, a subsample (N 40) of the previous sample (N 142) answered the exact same questionnaire after a three-month time period. All correlations between responses from Time 1 and 2 were above .6 and significant for each item. All correlations between the factors at Time 1 and Time 2 were also significant and higher than the recommended level of .7 (Nunnally 1978), aesthetic pleasure (.99) and novelty (.97).

5 Discussion

The exploratory factor analysis showed a good structure among the items constructing the scale. A confirmatory factor analysis is recommended in order to have a more robust validation of the scale. This instrument was created from the local respondents' language, so, although it allowed a much better understanding of the local context and proved to be a valid scale, the scale can only be generalizable inside the local context (Medellín, Colombia). Therefore, a validation of the instrument among different Spanish speakers outside the original context would be a next step. Also, this scale's input was completely visual oriented information, so its performance for other senses should be tested before use.

The results show an adequate fit of the different proposed items. All variables loads were above 0.6 for the factor aesthetic pleasure, and the correlation matrix showed high scores but no redundancy. *Llamativo* (eye-catching) had the lowest correlations inside the group of items, but it had a good theoretical conceptual fit and was still inside the range of desired values, so all of the items measuring aesthetic pleasure were retained. The difference in the correlations and the factor loadings of items measuring "aesthetic pleasure" and the items measuring the determinant "novelty" was clear after PCA and EFA. This is consistent with previous studies (Blijlevens et al. 2017; Marty et al. 2003)

and illustrates the importance of differentiating this construct from its determinants when selecting a proper measure. Consequently, five items are proposed to assess aesthetic pleasure in Spanish: *bonito* (beautiful), *hermoso* (gorgeous), *agradable* (nice/pleasing), *llamativo* (eye-catching) and *me gusta* (I like it). Although the final scale consists of five items, researchers could decide if they use a smaller number of items when considered necessary. Content validity should always be considered beforehand.

The adopted method allowed the research team to overcome the adaptation challenges by basing the construction of the instrument both on the respondent's relevant vocabulary and on the underlying aesthetic pleasure theory. The identification of the scale's core items through the elicitation technique allowed the team to gain insight into how aesthetic pleasure is understood and therefore expressed by local respondents. This resulted in a highly relevant initial pool of items, that were then filtered according to previously studied theories, eliminating the initial lack of understanding of the scale's items and their role in product evaluation. The use of the clustering method clarified the connections between items which resulted in a better understanding in the perceived similarities and differences between them. This allowed the research team to have a better picture on the construct's content. The final scale is considered equivalent to the original referenced instrument, as it can be used to measure the same construct under the same methodological considerations. Both instruments were validated inside the realm of product design proving to be reliable and valid instruments.

References

- Antmann, G., Ares, G., Varela, P., Salvador, A., Coste, B., Fiszman, S.: Consumers' texture vocabulary: results from a free listing study in three Spanish-speaking countries. *Food Qual. Prefer.* **22**(1), 165–172 (2011). <https://doi.org/10.1016/j.foodqual.2010.09.007>
- Augustin, M., Wagemans, J., Carbon, C.: All is beautiful? Generality vs. specificity of word usage in visual aesthetics. *Acta Psychologica* **139**(1), 187–201 (2012)
- Bartlett, M.S.: Tests of significance in factor analysis. *Br. J. Psychol.* **3**(Part II), 77–85 (1950)
- Behling, O., Law, K.: *Translating Questionnaires and Other Research Instruments* (2000). <https://doi.org/10.4135/9781412986373>
- Berghman, M., Hekkert, P.: Towards a unified model of aesthetic pleasure in design. *New Ideas Psychol.* **47**, 136–144 (2017). Elsevier
- Berlyne, D.E.: *Aesthetics and Psychobiology*. Hemisphere, Washington, DC (1971)
- Blijlevens, J., Thurgood, C., Hekkert, P., Chen, L.-L., Leder, H., Whitfield, T.W.A.: The aesthetic pleasure in design scale: the development of a scale to measure aesthetic pleasure for designed artifacts. *Psychol. Aesthetics Creativity Arts* **11**(1), 86–98 (2017)
- Borgatti, S.P.: Elicitation techniques for cultural domain analysis. In: Schensul, J.J., LeCompte, M.D., Nastasi, B.K., Borgatti, S.P. (eds.) *Enhanced Ethnographic Methods: Audiovisual Techniques, Focused Group Interviews, and Elicitation Techniques*, Ethnographer's toolkit, pp. 115–151, vol. 3. AltaMira, Walnut Creek (1999)
- Brislin, R.W.: Back-translation for cross-cultural research. *J. Cross Cult. Psychol.* **1**, 185–216 (1970)

- Capra, M.G.: Factor analysis of card sort data: an alternative to hierarchical cluster analysis. In: Proceedings of the "Human Factors and Ergonomics Society, 49th Annual Meeting, Orlando, Florida, USA (2005)
- Cha, E., Kim, K., Erlen, J.: Translation of scales in cross-cultural research: issues and techniques. *J. Adv. Nurs.* **58**(4), 386–395 (2007). <https://doi.org/10.1111/j.1365-2648.2007.04242.x>
- Clark, L.A., Watson, D.: Constructing validity: basic issues in objective scale development. *Psychol. Assess.* **7**(3), 309–319 (1995). <https://doi.org/10.1037/1040-3590.7.3.309>
- Deutcher, I.: Asking questions: linguistic comparability. In: Warwick, D., Osherson, S. (eds.) *Comparative research methods*, pp. 163–186. Prentice Hall, Englewood Cliffs (1973)
- DeVellis, R.F.: *Scale Development: Theory and Applications*, 2nd edn. Sage Publications, Newbury Park (2003)
- DeVos, G.: *Socialization for Achievement*. University of California Press, Berkeley (1973)
- Dutton, D.: *The Art Instinct*. Oxford University Press, New York (2009)
- Eremenco, S., Cella, D.S., Arnold, B.: A comprehensive method for the translation and cross-cultural validation of health status questionnaires. *Eval. Health Prof.* **28**(2), 212–232 (2005)
- Erol, T.: Dimensions of holistic automotive seat comfort experience: a card sorting approach. In: Proceedings of "The Human Factors and Ergonomics Society" Annual Meeting, vol. 62, no. 1, pp. 1007–1011 (2018). <https://doi.org/10.1177/1541931218621232>
- Eysenck, H.: The general factor in aesthetic judgements. *Br. J. Psychol. Gen. Sect.* **31**(1), 94–102 (1940). <https://doi.org/10.1111/j.2044-8295.1940.tb00977.x>
- Faerber, S., Leder, H., Gerger, G., Carbon, C.: Priming semantic concepts affects the dynamics of aesthetic appreciation. *Acta Physiol.* **135**(2), 191–200 (2010). <https://doi.org/10.1016/j.actpsy.2010.06.006>
- Fechner, G., Höge, H.: Various attempts to establish a basic form of beauty: experimental aesthetics, golden section, and square. *Empirical Stud. Arts* **15**(2), 115–130 (1997). <https://doi.org/10.2190/djyk-98b8-63kr-kudn>
- Field, A.: *Discovering Statistics Using SPSS*, 3rd edn. Sage Publications Ltd., London (2009)
- Field, A., Miles, J., Field, Z.: *Discovering Statistics Using R*. Sage, London (2014)
- Goldman, Alan H.: Aesthetic qualities and aesthetic value. *J. Philos.* **87**(1), 23–37 (1990)
- González-Calvo, J., González, V., Lorig, K.: Cultural diversity issues in the development of valid and reliable measures of health status. *Arthritis Care Res.* **10**(6), 448–456 (1997). <https://doi.org/10.1002/art.1790100613>
- Graf, L., Landwehr, J.: Aesthetic pleasure versus aesthetic interest: the two routes to aesthetic liking. *Front. Psychol.* **8**. <https://doi.org/10.3389/fpsyg.2017.00015>
- Hair Junior, J.F., Black, W.C., Babin, N.J., Anderson, R.E., Tatham, R.L.: *Análise multivariada de dados*, 6th edn. Bookman, São Paulo (2009)
- Halverson, S.L.: The cognitive basis of translation universals. *Target* **15**(2), 197–241 (2003)
- Hassenzahl, M., Monk, A.: The inference of perceived usability from beauty. *Hum.-Comput. Interact.* **25**, 235–260 (2010). <https://doi.org/10.1080/07370024.2010.500139>
- Hekkert, P.: Aesthetic responses to design: a battle of impulses. In: *The Cambridge Handbook of the Psychology of Aesthetics and the Arts*, pp. 277–299 (2015). <https://doi.org/10.1017/cbo9781139207058.015>
- Hekkert, P., Leder, H.: Product aesthetics. In: Schifferstein, H.N.J., Hekkert, P. (eds.) *Product Experience*, pp. 259–285. Elsevier, San Diego (2008)
- Hernández Belver, M.: *Psicología del arte y criterio estético*. Amarú, México (1989)
- Hilton, A., Skrutkowski, M.: Translating instruments into other languages: development and testing processed. *Cancer Nurs.* **25**(1), 1–7 (2002)
- Hines, A.M.: Linking qualitative and quantitative methods in cross-cultural survey research: techniques from cognitive science. *Am. J. Commun. Psychol.* **1993**(21), 72946 (1993)

- Hinkin, T.R.: A review of scale development practices in the study of organizations. *J. Manag.* **21** (5), 967–988 (1995). <https://doi.org/10.1177/014920639502100509>
- Hung, W., Chen, L.: Effects of novelty and its dimensions on aesthetic preference in product design. *Int. J. Des.* **6**, 81–90 (2012)
- Jacobsen, T.: Beauty and the brain: culture, history and individual differences in aesthetic appreciation. *J. Anat.* **216**(2), 184–191 (2010). <https://doi.org/10.1111/j.1469-7580.2009.01164.x>
- Jacobsen, T., Buchta, K., Köhler, M., Schröger, E.: The primacy of beauty in judging the aesthetics of objects. *Psychol. Rep.* **94**(Suppl. 3), 1253–1260 (2004). <https://doi.org/10.2466/pr0.94.3c.1253-1260>
- Jöreskog, K.G., Sörbom, D.: *Advances in Factor Analysis and Structural Equation Models*. Abt Books, Cambridge (1979)
- Kaiser, H.F.: The application of electronic computers to factor analysis. *Educ. Psychol. Meas.* **20**, 141–151 (1960)
- Kapuscinski, A.N., Masters, K.S.: The current status of measures of spirituality: a critical review of scale development. *Psychol. Relig. Spirituality* **2**(4), 191–205 (2010). <https://doi.org/10.1037/a0020498>
- Kinzie, J.D., Manson, S.M., Vinh, D.T., Tolan, N.T., Anh, B., Pho, T.N.: Development and validation of a Vietnamese-language depression rating scale. *Am. J. Psychiatry* **139**(10), 1276–1281 (1982)
- Maneesriwongul, W., Dixon, J.K.: Instrument translation process: a methods review. *J. Adv. Nurs.* **48**(2), 175–186 (2004). <https://doi.org/10.1111/j.1365-2648.2004.03185.x>
- Marsella, A.J., Hirschfield, R.M., Katz, M. (eds.): *Measurement of Depression*. Guilford, New York (1987)
- Marsella, A.J., Kinzie, D., Gordon, P., et al.: Ethnocultural variations in expression of depression. *J. Cross Cult. Psychol.* **4**, 435–458 (1973)
- Martin, J.: Beyond exchange: APPRAISAL systems in English. In: Hunston, S., Thompson, G. (eds.) *Evaluation in Text: Authorial Stance and the Construction of Discourse*, pp. 142–175. Oxford University Press, Oxford (2000)
- Martindale, C., Moore, K., Borkum, J.: Aesthetic preference: Anomalous findings for Berlyne's psychobiological theory. *Am. J. Psychol.* **103**, 53–80 (1990). <https://doi.org/10.2307/1423259>
- Marty, G., Conde, C.J.C., Munar, E., Rosselló, J., Roca, M., Escudero, J.T.: Dimensiones factoriales de la experiencia estética [Factorial dimension of aesthetic experience]. *Psicothema* **15**(3), 478–483 (2003)
- McDermott, M.A., Palchanes, K.: A process for translating and testing a quantitative measure for cross-cultural nursing research. *J. New York State Nurs. Assoc.* **23**, 12–15 (1992)
- Mechanic, D.: Medical sociology: some tensions among theory, method, and substance. *J. Health Soc. Behav.* **30**, 147–160 (1989)
- Montgomery, S.L.: Translation of scientific and medical texts. In: *Encyclopedia of Language & Linguistics*, vol. 13. Elsevier, Amsterdam (2006)
- Nunnally, J.C.: *Psychometric Theory*, 2nd edn. McGraw-Hill, New York (1978)
- Page, C., Herr, P.M.: An investigation of the processes by which product design and brand strength interact to determine initial affective and quality judgments. *J. Consum. Psychol.* **12**, 133–147 (2002). https://doi.org/10.1207/S15327663JCP1202_06
- Pett, M.A., Lackey, N.R., Sullivan, J.J.: *Making Sense of Factor Analysis: The Use of Factor Analysis for Instrument Development in Health Care Research*. Sage Publications Inc., Thousand Oaks (2003)
- Roussos, L., Dentsoras, A.: Formulation and use of criteria for the evaluation of aesthetic attributes of products in engineering design. In: *International Conference on Engineering Design, ICED 2013*, pp. 1–10 (2013)

- Shortess, G., Clarke, J., Shannon, K.: The shape of things: but not the golden section. *Empirical Stud. Arts* **15**(2), 165–176 (1997). <https://doi.org/10.2190/3ubv-2lju-t2t9-d7xn>
- Sinha, R., Boutelle, J.: Rapid information architecture prototyping. In: *Proceedings 2004 Conference on Designing Interactive Systems Processes, Practices, Methods, and Techniques - DIS 2004* (2004). <https://doi.org/10.1145/1013115.1013177>
- Smith, J.J.: Using ANTHROPAC 3.5 and a spreadsheet to compute a free list salience index. *Cogn. Technol. Work* **1**, 179–196 (1999)
- Spencer, D., Garrett, J.: Card Sorting, pp. 82–85. Rosenfeld Media, Brooklyn (2009)
- Sperber, A.D., DeVellis, R.F., Boehlecke, B.: Cross-cultural translation. *J. Cross Cult. Psychol.* **25**, 501–524 (1994)
- Stewart, M.: *The Spanish Language Today*. Routledge, London (2003)
- Thompson, B.: *Exploratory and Confirmatory Factor Analysis: Understanding Concepts and Applications*. American Psychological Association, Washington, DC (2004)
- Wang, W., Lee, H., Fetzer, S.: Challenges and strategies of instrument translation. *West J. Nurs. Res.* **28**, 310–321 (2006)
- Sequeiros, X.R.: Translation: pragmatics. In: *Encyclopedia of Language & Linguistics*, vol. 13. Elsevier, Amsterdam (2006)