

Experimental evaluation of Ellis scale for assessment of product's industrial design

*Andrei Dumitrescu** and *Mihail Purcărea*

POLITEHNICA University of Bucharest, Department of Manufacturing Engineering, Splaiul Independentei 313, Bucharest, Romania

Abstract. The product development team has the task, amongst many others, to conceive the product's industrial design. Because it was proven and acknowledged that product's industrial design plays an important role in its success in the marketplace, the proposed design has to be evaluated based on an objective method. S.R. Ellis has proposed a method with seven criteria. The method was used by some researchers, but it has never been properly evaluated by an independent party. This paper presents the results of a scientific experiment aimed to test Ellis' method.

1 Introduction

There is a general belief that the designer is just conceiving the product aesthetics according to her/his free will. And she/he is acting on inspiration and from idea to detailed project is just a linear array of activities. The truth is quite different. The designer has a structured approach, not like as the engineer has, but structured nonetheless. Also, most steps of the development process are iterative.

Part of the development process are the assessment phases. One such a phase is the assessment of the generated concepts in order to select the best one and to continue its development. Other phases are focused on the assessment of the designed product at certain levels of development to determine whether the design process is worthwhile or if significant changes to the design product were required. The emphasis of assessment is on revealing the positive characteristics to be maintained and the negative ones to be eliminated or at least mitigated.

To succeed in its endeavor, the designer should apply an objective method of assessment. Besides specific activities to be performed and specific conditions to be ensured, any assessment should use a set of clear and objective criteria. There are several attempts in the specialized literature to establish an objective and easy-to-use method. In the Treaty of Design [1], several assessment methods are presented and analyzed. Some methods are using more or less the same criteria, like balance, harmony or proportion. For example, a product is generally considered to possess a good proportion if it had a ratio of 1.618 between two overall dimensions (the so-called Golden Section). This precise ratio can be obtained algebraically from the Fibonacci series or geometrically from a square and a semicircle. An experiment [2] proved that the ratio was a correct measure for proportion.

*Corresponding author: andrei.dumitrescu@upb.ro

A challenge to the objectivity of assessment is represented by the empirical assumptions that are frequently encountered in industrial design literature. For example, some theoreticians of the beginning of the twentieth century postulated that certain geometric figures are more beautiful when filled with a certain color (square – red; circle – blue; etc.). Several researches [3] proved that this assumption was wrong. This sort of empirical assumption should be avoided in the name of objectivity.

Arguably the most used method for assessment of industrial design is the Ellis scale [4]. A quick search on Google Scholar performed while writing this paper indicated 23 citations for Ellis' dissertation where the method was proposed. For example, Blijlevens, Creusen, and Schoormans [5] referred to Ellis' criteria when they attempted to introduce a new assessment method. Brunel and Kumar [6] and Koprda [7] actually used Ellis' criteria in the assessment of certain products.

2 Ellis scale for industrial design assessment

Seth Robert Ellis [4] proposed a scale for assessing consumer perceptions of the aesthetic dimensions of durable goods. The American author also intended to use the scale in segmenting the market.

Amongst the many criteria used in industrial design literature, Ellis chose the following: a) simplicity/complexity; b) harmony; c) balance; d) dynamics; e) unity; f) timeliness and style and g) novelty. According to Ellis, the definitions of the above criteria are as follows.

Simplicity expresses the amount of information (visual, etc.) provided by the product design. The less information is provided, the simpler the product is. The richer the information, the more complex the product is. **Harmony** refers to the degree of experienced similarity or concordance among the various parts of a product's visual design, with respect to elements like shape, size and color. **Balance** refers to how the arrangement and structure of the parts of an object are interpreted. **Dynamism** refers to how the parts and details of the product are arranged in order to create the impression of movement or tension. **Unity** indicates how the parts of the product are integrated into one single unit. **Timeliness / style** refers to the degree to which product aesthetics is integrated in the contemporary era and adapts itself to a modern style. **Novelty** expresses the innovative character of the product. [4]

Ellis proposed a rating scale for assessment of industrial design inspired by the Osgood's semantic differential. Instead of the usual scale [-2 -1 0 +1 +2] for semantic differential, Ellis introduced a 5 point Likert scale. For each criteria, Ellis proposed three-four pairs of adjectives for each criteria. Each pair had a "positive" left-hand element and a "negative" right-hand element, with the notable exception of the first criterion: simplicity. It was hard to evaluate which was positive: simplicity or complexity.

Ellis studied pairs of adjectives in terms of discrimination and reliability. Except for some minor issues, he concluded that all pairs of adjectives are reliable and allow a discrimination between the designs of the different products. It should be noted that he found that the "simple - complex" pair does not allow proper discrimination [4, p. 168]

3 Design of experiment

Ellis did analyzed his method and found it reliable and discriminatory. He discovered some minor shortcomings, but overall the method functioned well. But even some experimenters used it, the Ellis scale was not tested and validated by other researchers. So, the authors decided to put this method under scrutiny. They considered that the method should be analyzed as an assessment method aimed to be used by any person (expert or layman).

The research objectives were established:

1. Does the Ellis scale accurately indicate the aesthetic impression left by a product?
2. Does the Ellis scale reveal the real aesthetic features of the analyzed product?
3. Does the method allow the identification of products with outstanding design?
4. How clear are the proposed Ellis criteria / adjectives?
5. Which adjective pair (from those proposed by Ellis) should be used for each criterion?

In order to answer the research questions, the authors decided to ask experiment participants to aesthetically assess a series of products from the same class, using the Ellis scale plus an additional assessment. The additional assessment was an overall evaluation of product aesthetics (“Please assess the overall design of the product.”). The additional assessment had associated a Likert scale (“Excellent 1 ... 5 Totally inappropriate”). The purpose was to compare, for each participant, the rating of overall assessment of each product to the mean of ratings of Ellis criteria. If the overall rating was equal or close to the mean of Ellis ratings, then the answer to the first research question would be affirmative.

The second research question would be answered by comparing the aesthetic features determined by the authors by direct evaluation of the products to the aesthetic features revealed using the Ellis scale. It was assumed that all Ellis criteria had equal weight. The answer to the third research question would be found by comparing the means of mundane products to those of outstanding ones in terms of product aesthetics. The clarity of criteria or adjectives would be indicated by the efficiency of discrimination in assessment (fourth research question). The answer to the last research question would be found by simply asking the participants to choose the right pair of adjectives according to their own opinion. Because the scale was aimed to be used also by non-professionals, it was important to evaluate how easy would be to use a pair like “well resolved ... poorly resolved”.

Some research objectives needed the assessment of products from the same class and it was decided to use images of office desks. In the first phase, there were selected 40 images of desks. In the second phase, it was decided to have two images of different traditional desks (traditional, but not anonymous) and two different modernistic desks. After several iterations, there were selected the images displayed in Figures 1 – 4. The following set of adjective pairs was used for the assessment of the four products: a) simple – complex; b) harmonious – disharmonious; c) balanced – not balanced; d) dynamic – static; e) integrated – disintegrated; f) contemporary – out of date; g) innovative – common. As much as possible, there were used the pairs that Ellis recommended, but the authors changed two adjectives, respectively the opposites of balanced (“not balanced”) and integrated (“disintegrated”).



Fig. 1. Desk 1.



Fig. 2. Desk 2.



Fig. 3. Desk 3.



Fig. 4. Desk 4.

4 Analysis of experimental results

The experiment was carried-out with 320 participants (189 female and 131 male participants). All participants were students enrolled at a large technical university in Romania. The list of adjective pairs and the rating scales had been translated into Romanian prior to the experiment. The product images were presented on computer displays of the same model. The whole experiment duration was four months and all sessions were supervised by the first author.

The first step was the assessment of aesthetic significance made by authors using the Ellis adjectives. The products were found to be: 1 - balanced, static, integrated; 2 - simple, static, old-fashioned; 3 - complex, harmonious, contemporary, innovative; 4 - complex, integrated, contemporary, innovative.

The results of participants' selection of the most significant pairs of opposite adjectives (in their perception) are displayed in Table 1. Because of the paper's formatting constraints, only the highest percentage is presented in the table. Only the "Harmony" criterion had a winning pair, the other criteria had the highest ranked pair at around 50% or less. That meant they were less significant.

Table 1. Participants' selection of pairs of opposite adjectives.

Criterion	Pair of opposite adjectives	Selected [%]
Simplicity	simple ... complex	48.68
Harmony	harmonious form ... disharmonious form	64.55
Balance	a controlled sense of balance ... no sense of balance	37.04
Dynamics	energetic ... stagnant	49.74
Unity	a well-organized layout ... poorly organized layout	37.04
Timeliness and style	modern design ... traditional design	51.32
Novelty	innovative design ... a routine design	47.19

The participants' assessment of products' aesthetics using Ellis scale were recorded and the means for each criterion are displayed in Table 2. Also in Table 2, there are presented the mean of all criteria ratings (penultimate row – as mean of the first seven rows) and the mean of the direct overall assessment made by participants. It was assumed that all Ellis criteria had equal weight. This assumption should be changed in future researches and different weights should be assigned to different criteria.

It was obvious that a mean value around 3 (actually in the interval 2.5 – 3.5) lacked practically a proper significance for the respective criterion. The significant values are marked with bold and italics. There were considered important the tendencies and less the absolute values. It was worth noting that: simplicity had 3 out of 4 significant values; harmony – 1; balance – 1; dynamics – 2; unity – 1; timeliness – 2 and novelty – 3. That meant that only simplicity and novelty are relevant criteria.

Table 2. Assessment of products using Ellis scale and directly by participants.

	Product 1	Product 2	Product 3	Product 4
Simplicity	2.56	2.20	4.11	4.13
Harmony	2.50	2.90	2.36	2.71
Balance	2.11	2.68	2.94	2.66
Dynamics	3.60	3.59	2.58	2.79
Unity	2.35	2.89	2.62	2.36
Timeliness and style	2.93	3.28	1.96	2.21
Novelty	3.49	3.69	1.75	1.88
Mean of Ellis ratings	2.79	3.03	2.62	2.68
Mean of overall assessment	2.83	3.08	2.07	2.51

Another relevant aspect in Table 2 was the relative small difference between the mean of Ellis ratings and the direct overall assessment made by participants, with the notable exception of product 3 (product with an outstanding design). Did this mean that the Ellis scale accurately indicate the aesthetic impression left by a product? The answer appeared to be affirmative, but a t-test was performed to validate this apparent conclusion.

The null hypothesis (H_0) was: the rating of direct assessment = the mean of Ellis ratings. Applying the t-test for product 1, it was found that mean of overall assessment ($M = 2.83$, $SD = 0.75$) was equal to the mean of Ellis ratings ($M = 2.79$, $SD = 0.25$), $t_{stat} = 0.63 < 1.64 = t_{cr}$, but $p = 0.26 > 0.05$, so the result was inconclusive. It was a similar situation for product 2: the mean of overall assessment ($M = 3.08$, $SD = 1.07$) against the mean Ellis ratings ($M = 3.03$, $SD = 0.47$), $t_{stat} = 0.54 < 1.64 = t_{cr}$, but $p = 0.29 > 0.05$, so the result was also inconclusive. In case of product 3, the mean of overall assessment ($M = 2.07$, $SD = 1.08$) against the mean Ellis ratings ($M = 2.62$, $SD = 0.37$), $t_{stat} = -8.1 < 1.64 = t_{cr}$, so the null hypothesis was rejected. In case of product 4, the mean of overall assessment ($M = 2.51$, $SD = 1.63$) against the mean Ellis ratings ($M = 2.68$, $SD = 0.51$), $t_{stat} = -2.1 < 1.64 = t_{cr}$, so the null hypothesis was rejected. With two rejections and two inconclusive results, it should be concluded that Ellis scale did not reflect the participants' aesthetic impression.

Considering the significant values from Table 2 (marked with bold and italic), there were identified the corresponding features (adjectives) and were recorded in Table 3. The identified features were compared with those resulted from authors' assessment (see beginning of "Analysis of experimental results"). It was noted the match in case of all products, so the Ellis scale allowed the identification of real aesthetic features.

Remembering that the positive adjectives are on the left side of the Ellis scale associated with mark "1", it was clear that products with a low mean of Ellis ratings were considered more beautiful than products with a higher mean.

Table 3. Confirmation of significances resulted from Ellis scale.

Product	Significance according to Ellis ratings	Confirmation against authors' assessment
Product 1	balanced, static, integrated	Checked
Product 2	simple, static, out of date	Checked
Product 3	complex, harmonious, contemporary, innovative	Checked
Product 4	complex, integrated, contemporary, innovative	Checked

Ordering the products according to the mean of Ellis ratings, it was found that the order is correct, product 3 being the most aesthetically pleasing and product 2 – the least with a difference of 0.41 between them. But the mean of overall assessment despite indicating the same order was presenting a greater difference – 1.01. So, the Ellis scale allowed the identification of outstanding design, but not as efficient as direct assessment.

5 Conclusions

Based on the analysis of experimental results, the conclusions of the experiment were:

- The Ellis scale did not accurately indicate the aesthetic impression left by a product.
- The Ellis scale revealed correctly the real aesthetic features of the analyzed products.
- The method allowed the identification of the product with outstanding design, but the direct assessment proved to be more efficient.
- From the criteria proposed by Ellis, simplicity and novelty proved to be the most efficient, while balance and harmony needed replacement or better explanations.
- Only the “harmonious - disharmonious” pair of adjectives was chosen by a significant number of participants.

References

1. A. Dumitrescu, *Treaty of Industrial Design* (in Romanian) (Romanian Academy Publishing House, Bucharest, 2013)
2. A. Dumitrescu, Politehn. Univ. Bucharest Sci. Bull. Ser. A Appl. Math. Phys. **71**(1), 33-42 (2009)
3. A. Dumitrescu, *New Researches Regarding Relationship between Elementary Geometric Shapes and Basic Colours*, Annals of DAAAM Proceedings, 1041-1042 (2011)
4. S.R. Ellis, *A psychometric investigation of a scale for the evaluation of the aesthetic element in consumer durable goods* (University of Arizona, 1993)
5. J. Blijlevens, M.E.H. Creusen, & J.P.L. Schoormans, *IJDesign* **3**(3), 27-35 (2009)
6. F.F. Brunel, & R. Kumar, *Design and the Big Five: Linking Product Aesthetics to Product Personality*, NA - Advances in Consumer Research **34**, 238-239 (2007)
7. T. Koprda, *The influence of packaging design attractiveness on consumer's behaviour*, “Current trends and challenges in economics and management” Conference in Poprad, 138-145 (2016)