Chapter V Discussion

Many attempts have been made in the past to measure aesthetic sensitivity. However, these measures either show poor psychometric properties, were developed for specific experimental settings, are rather time-consuming, or focus exclusively on art works. The research presented here describes the development of scales to measure aesthetic sensitivity that avoid the pitfalls of previous approaches. Additionally, the present research highlights and details specific methodological issues and shows how existing methods, such as multidimensional unfolding and conjoint analysis, can be used to effectively deal with issues of scale development.

Chapter 1 described a general approach to scale development that was used to develop a scale for measuring aesthetic sensitivity. Aesthetic sensitivity was defined in this chapter as an individual's perception of the beauty of objects in his or her immediate environment. Because the research focus was on the immediate environment the stimuli used for scale development were objects taken from people's everyday environment such as cutlery, vehicles, furniture, and jewellery. Even though Gestalt Psychology and Berlyne's new experimental aesthetics discovered aesthetic principles that might be important for the aesthetic appeal of visual objects, such as symmetry, balance, clarity, color, novelty, no such characteristics were available for everyday objects. It was thus difficult to determine a priori which stimuli would induce a positive response in an aesthetically sensitive perceiver. However, in order to develop a scale for measuring aesthetic sensitivity, stimuli were needed that differ in their aesthetic values. Consequently, the stimuli chosen for scale development were *obvious* exemplars of ugly and beautiful objects. For evaluating the factor structure of the scale a study was conducted with a North-American sample. Results of an exploratory factor analysis (EFA) showed that the scale consists of two rather independent factors, one representing the ugly stimuli, thus labelled "ugly," and the other representing the beautiful objects, thus labelled "beauty". This result suggests that aesthetic sensitivity for ugly objects is different from sensitivity for beautiful objects and that a person who is sensitive to the aesthetic value of ugly objects is not necessarily sensitive to the aesthetic value of beautiful objects, and vice versa. Results further showed that the scales are reliable and give initial evidence for convergence and divergent validity of the measure. In a study conducted with a German student sample, higher correlations with measures of convergent and divergent validity were found for the beauty scale than for the ugly scale. This result was interpreted as suggesting that the beauty scale focuses on the perception of aesthetically pleasing objects, as

do other instruments developed for measuring aesthetic sensitivity. In contrast, the ugly scale appears to measure an independent dimension of aesthetic sensitivity rather than the opposite of the aesthetic sensitivity for beautiful objects. Consequently, the ugly scale did not show good convergent validity with existing measures of aesthetic sensitivity, which focus on perceptions of beauty.

The results of the scale development described in Chapter 1 revealed some problems with the chosen approach. Specifically, two major problems appeared because empirical aesthetics mostly uses visual stimuli and researchers have very little knowledge about relevant judgment criteria for these stimuli. One problem was that the heterogeneity of the visual stimuli used for scale development resulted in relatively low goodness of fit measures in Study 1 (EFA) and Study 2 (CFA), as compared, for example, with scales using verbal stimuli. In Study 1, the amount of variance explained by the two factors was rather small. It was argued that, by nature, visual stimuli are much more heterogeneous from each other than verbal items. Each visual stimulus varies in a variety of aspects such as form, color, texture and much more; whereas verbal statements can be constructed rather homogenous. The fact that communalities - and consequently the amount of variance explained by the factors - tend to be rather low for scales with heterogeneous stimuli can explain the low amount of variance explained by the EFA in Study 1. The relatively low goodness of fit measures in CFA show that a significant amount of variance remains in the data even when taking into account the variance explained by the two factors identified by the EFA.

A second problem concerns the validity of the new measure. The first scale (described in Chapter 1) was developed using the classic approach of scale development. The classic approach allows selecting stimuli using statistical measures such as the inter-item correlation or factor loadings. Technically, there is no need for understanding which characteristics of the stimuli relate to, and thus help assessing the construct. This might not be considered problematic when well-known constructs are the focus of research. In this case enough knowledge about which stimuli to use for assessing the construct is available and other reliable and valid scales can be used for assessing the validity of a new measure. However, it is different in a situation when the aim of scale development is to assess a construct about which relatively little knowledge exists, such as aesthetic sensitivity. One difficulty encountered when developing the ugly and beauty scale (UgBeaScale) was thus that few or no other measures with good psychometric properties exist that could be used to assess the validity of the new scale. In the case of the beauty scale, the new measure showed significant

correlations with other measures, so that the beauty scale could be considered assessing the aesthetic sensitivity construct. However, the effect sizes of the correlations found with other scales suggests that both the beauty and the ugly scale have a similar problem, namely that their validity is doubtful. Even though most of the correlations were significant, they were overall very low in terms of effect size ranging from r = .13 to r =.29. The ugly scale did not show significant correlations with most of the existing measures, making it difficult to understand what it measures. One reason for a lack of good validity of both the ugly and the beauty scale can be seen in the measures used for assessing the validity. The new scale was developed for measuring aesthetic sensitivity for *everyday objects.* Other existing measures rather focus on different aspects of aesthetic sensitivity, mostly on the aesthetic sensitivity for art work.

Another difficulty appears if it is unkown which characteristics of the used stimuli are relevant for assessing the construct. Stimuli might then represent different characteristics of which some might be relevant for assessing the construct whereas other might not. This non-systematic variation between stimuli creates problems for the validity of the scale. In the worst case the stimuli are so heterogeneous that the respondents are not able to systematically judge the stimuli according to a certain criteria (such as their aesthetic value). In this case, responses to the stimuli might be given rather randomly and the final data might reflect nothing more than guessing. However, even when dealing with rather homogenous visual stimuli, the validity of an instrument using these stimuli is difficult to assess because it is unclear which aspects of the stimuli respondents use for their judgment. Because no characteristics of the (visual) stimuli used for scale development in the research described in Chapter 1 are known and because no knowledge is gained by the used process of scale development about the characteristics, it remains rather unclear what construct the scale measures. This specifically applies to the ugly scale.

The results of the research described in Chapter 1 illustrated that in order to construct a scale that measures aesthetic sensitivity for everyday objects a different approach to scale development has to be taken. This new approach needs to take into consideration that aesthetic sensitivity for everyday objects has to be assessed using visual stimuli but that not enough knowledge about these stimuli is available to select a rather homogeneous pool of stimuli for scale development. Consequently, it was necessary to acquire more knowledge about the visual stimuli to be included in a new measure of aesthetic sensitivity and their characteristics as they related to the construct of aesthetic sensitivity. The aim of the research

described in Chapter 2 was therefore to identify such characteristics. Additionally, because existing measures of aesthetic sensitivity focus on beautiful objects (such as works of art), ugly stimuli were not used in the studies described in Chapter 2.

The research described in Chapter 2 was conducted to address some of the problems encountered during the scale development presented in Chapter 1. In order to understand more about the visual stimuli and how their characteristics relate to aesthetic sensitivity, the aim of the research reported in Chapter 2 was to identify important dimensions that are used as common judgment criteria for different classes of everyday objects. These dimensions were then used to create systematically varied, real-life stimuli. Studies 4 and 5 were also used to identify characteristics of everyday objects that are relevant when aesthetic judgments about the objects are made. The logic here was that if relevant judgment dimensions can be identified, they can be used to construct stimuli that differ only in these relevant judgment dimensions. These stimuli can then be used for constructing a new scale measuring aesthetic sensitivity. Identifying relevant aesthetic judgment dimensions and using them for constructing systematically varied, visual stimuli would help dealing with the two main problems identified in the previous scale development (Chapter 1). As it concerns the validity problem, the patterns of results found when using such a scale could only reflect individual differences in the ability to perceive the differences across the stimuli on the relevant judgment dimensions. Thus, the scale should have high content validity because it uses systematically varied stimuli. Moreover, stimuli constructed in this way represent rather homogenous stimuli to create larger communalities across stimuli and thus more variance explained in a factor analysis.

For the research presented in Chapter 2, everyday objects of four different object classes were chosen as stimuli - namely paintings, offices, car interiors, and cutlery. Aesthetic dimensions of the different classes of everyday objects were identified using expert interviews (Study 4) and multidimensional unfolding (Study 5). In the expert interviews students of art history, interior designers, object-oriented designers and architects were asked to identify the aesthetic properties they used to judge the aesthetic value of each stimulus. The resulting aesthetic properties were summarized in categories for each object class. In Study 5, multidimensional unfolding was used to identify whether non-experts use the aesthetic properties identified in the expert interviews as common criteria for evaluating the aesthetic of everyday objects. The categories derived from the expert interviews in Study 4 were used in Study 5 for labelling the empirically derived aesthetic dimensions. For the different object

classes either two or three different dimensions were identified by multidimensional unfolding. The dimensions of the obtained solutions were labelled as follows: for paintings, the three dimensions were labelled “shape”, “painting technique,” and “contrast”; for the office stimuli, the two dimensions were labelled “arrangement” and “atmosphere”; for car interiors, the three dimensions were labelled “simplicity, “clarity,” and “technology”; and, finally, for the cutlery stimuli, the dimensions were labelled “proportion,” “harmony,” and “perceived ergonomics.” These dimensions represent common criteria used by non-experts for judging the aesthetic values of the given stimuli. The results of Study 4 and 5 indicate that *different* aesthetic judgment dimensions are relevant for each object class or in other words that each object class has its own idiosyncratic set of relevant characteristics or dimensions. When judging the aesthetic value of everyday objects people seem to use criteria that are specific to the object class to which an object belongs. These results are contrary to the assumption that people use the same general dimensions for evaluating the aesthetic values of everyday objects, as has been proposed in the concept of Gute Gestalt from Gestalt Psychology or Berlynes’ collative variables for other aesthetic objects. Consequently, when choosing stimuli for scale development, researchers need to identify judgment dimensions that are relevant for the specific object class. Only when these dimensions are known can they be used to choose stimuli that differ in their aesthetic values on these dimensions. In sum, Study 4 and 5 revealed important knowledge about the visual stimuli and their characteristics as they relate to aesthetic sensitivity. This knowledge was then used for scale development.

Instead of using the aesthetic judgment dimensions identified by multidimensional unfolding in Study 5 for *choosing* stimuli that vary on these dimensions from a larger pool of stimuli from a specific object class, they were used in Study 6 to *construct* visual stimuli that only vary on the relevant aesthetic judgment dimensions. The new sets of visual stimuli represent everyday objects of the four object classes that were used in Studies 4 and 5 - namely, paintings, offices, car interiors and cutlery. The stimulus within each object class that was judged in Study 5 as most aesthetic was used as template for constructing the remaining stimuli of that object class. Specifically, the template stimulus was varied systematically on the relevant aesthetic dimensions, rendering new stimuli that vary systematically on the factors (dimensions) that influence aesthetic judgments. Again, using MDU for data analysis, Study 6 examined whether the aesthetic dimensions derived from Study 5 could be successfully implemented in the new sets of stimuli. The results from Study 6 confirmed the dimensions used to create the new stimuli for all four object classes. Moreover, the results

suggested that the different dimensions are differently important for the overall aesthetic judgment. Obviously, the combination of different levels on the different dimensions can lead to different preference orders depending on the importance of a certain dimension in relation to others. This explanation is consistent with the pattern of results for some object classes in Study 6. The relative importance of the judgment dimensions for the overall aesthetic judgment was further investigated in Study 9 (Chapter 4, see below).

The studies described in Chapter 3 were designed to develop a scale measuring visual aesthetic sensitivity using the systematically varied stimuli that were constructed in Study 6 (Chapter 2). The process of scale development included stimulus reduction, reliability (Study 7) and validity (Study 8) testing. The initial stimulus pool representing each of the four different object classes (i.e., paintings, offices, car interiors, and cutlery) was reduced through pairwise comparisons within each object class. Using the data from the North American sample, the factorial structure and the reliability of the remaining 15 stimuli was explored. Three factors were identified. One factor represented the three office stimuli and the four car interior stimuli. The second factor represented the painting stimuli. The third factor represented the cutlery stimuli. The first factor was labelled “space” because the stimuli loading on it represented a room or space in which a person can move around. The second and third factors were called “painting” and “cutlery,” respectively, because these were the only objects they contained. The test-retest results showed that the scores received with the scale are reliable over time. Results of Study 8 conducted with a German student sample however did not succeed at providing evidence for the convergent and divergent validity of the scale. Reasons for the lack of both kinds of validity were identified in the process of stimulus reduction as well as in the chosen response format (i.e. rating scales). The data of Study 7 was collected in downtown Montreal, Quebec. In that situation, it seemed the most prudent approach to choose an easy and non-demanding assessment method, given the limited time that participants had when approached downtown. Consequently, participants were given 7-point Likert-type rating scales for judging the aesthetic values of the stimuli. For consistency reasons the same rating scales were used in Study 8, with the German student sample. This procedure was used with the expectation that all participants (both North Americans and Germans) would use the rating scales to differentiate not only between stimuli of different object classes but also between the aesthetic values of stimuli *within* each object class. The data analysis, however, showed that most participants did not take advantage of the available numeric values of the rating scale to differentiate between objects within an object class.

Instead, they evaluated the aesthetic values of all objects within an object class rather similarly, and only differentiated between the aesthetic values of different object classes. Consequently, future applications of the scale, and future investigations of its psychometric properties, should use a version of the scale that forces individuals to differentiate between the stimuli *within* an object class (not only across object classes), such as rank orderings or paired comparisons.

The results of the research described in Chapter 3 should be seen as indication for the fact that the use of rating scales is not always appropriate even though it might result in good goodness of fit measures for an instrument. At first glance, the data from Study 7 and 8 could have been interpreted as providing at least some evidence for the scale being a reliable and somewhat valid scale for measuring aesthetic sensitivity. For instance, in Study 7 the results of the exploratory factor analysis indicated a three factorial solution explaining over 54% of the variance, internal consistency suggested that the subscales and the scale overall are highly internally consistent and the correlation values of the test-retest reliability demonstrated stability of the construct. Study 8 showed good internal consistency and good test-retest reliability of the scale. Moreover, significant correlations were found for the aesthetic sensitivity scale with the response subscale of the CVPA measure and with the SOP scale. Several additional correlations were found for the subscales of the aesthetic sensitivity scale with some of the CVPA subscales, self-rating for aesthetic sensitivity and judgment certainty. Without an additional analysis of the data as described in the discussion of Chapter 3, these results could have been interpreted as initial evidence for the psychometric soundness of the scale. The lack of convergent and divergent validity might for example have been interpreted as a result of knowing too little about the aesthetic sensitivity construct and thus lacking adequate measures for assessing the validity of the present scale. Yet, when the data were examined more closely, it was found that the rating scales were only used to differentiate judgments between object classes, but not stimuli within an object class. Consequently, at least when very little knowledge about a construct is available, the use of rating scales for constructing a scale measuring this construct should be done with great care.

The research described in Chapter 4 focuses mainly on two aspects. One aspect revisits a claim that was made in Study 6 (Chapter 2). In that chapter, I argued that different dimensions of a stimulus might be differently important for the overall aesthetic judgment. Thus, instead of assuming that all relevant dimensions of a stimulus have the same impact on the overall judgment, the relative importance of each dimension that was used to construct a stimulus

was investigated using Conjoint Analysis in Study 9. A second aspect of Chapter 4 concerns that aesthetics research traditionally has used the deviation of an individual's aesthetic judgment from an external standard as indication of his/her aesthetic sensitivity for the aesthetic value of stimuli. Such approaches are thus measures of agreement rather than direct measures of a person's aesthetic sensitivity. In order to overcome this problem in the aesthetic sensitivity literature, the second aim of Study 9 was to establish an external standard that is independent of an average expert's or non-expert's judgment.

Using conjoint analysis in Study 9, the average relative importance of each dimension and the value for each variation of a dimension (i.e. the part-worth utility) and for each stimulus (i.e. the utility) were evaluated. As expected, the results showed that different aesthetic dimensions of objects in different classes were differently important for the overall aesthetic judgment. For instance, the most important dimension for car interiors appears to be "simplicity", the second important is "clarity" and the least important is "technology." That is, the dimensions are not equally important to judgments of car interiors. Study 9 thus revealed a precise picture of the relative importance of dimensions for evaluating the aesthetic value of the given stimuli. The calculation of the overall utility for each stimulus results in a specific rank order within each object class. These rank orders - called "optimal rank orders" can be used as an external standard for evaluating any individual's rank order of the stimuli. Thus, the rank order could be used to measure aesthetic sensitivity for this and future research. The deviation of any individual's rank order from the "optimal" rank order as established in Study 9 could be evaluated using a Spearman correlation. It could be concluded that the higher the correlation of an individual's rank order with the optimal rank order, the more aesthetically sensitive a person is because the person perceives more differences between stimuli (i.e., is more sensitive to the relative combinations of features within a stimulus). Because the stimuli used in this study were constructed systematically, differences between individuals in evaluating the aesthetic value of the stimuli can be interpreted as being due to relevant aesthetic dimensions inherent in the stimuli (i.e, the dimensions themselves change across stimuli within an object class), which individuals may be able to perceive or not. The external standard proposed here is therefore different from the commonly used criteria in that: (a) it is based on knowledge about the properties of stimuli gained from interviews with experts and from multidimensional unfolding studies with non-experts, (b) the relative importance of each aesthetic dimension on the aesthetic judgment is taken into account, and (c) it is not a measure

that is relative to a certain reference group (such as an average judgment) but rather based on the properties of the stimuli themselves.

With the presented analyses, all tools are available to investigate the reliability and validity of the developed visual aesthetic sensitivity scale for objects (VAESO). This scale consists of the 34 new stimuli constructed in Study 6. The development of the external standard that can be used to assess the aesthetic sensitivity with these stimuli was described in Study 9 of the present work.

Future research

The present research demonstrates the benefits of using existing statistical methods (multidimensional unfolding, conjoint analysis) for the construction of systematically varied stimuli. The insights gained in the present research are significant for future scale constructions in empirical aesthetics research. Given the step-by-step description of the scale development in Chapters 2 and 3, researchers who need certain visual stimuli may model their stimulus and scale construction after the procedure proposed and used here. Future research needs to continue and expand this approach. The present work demonstrated, for example, how the combination of methods such as expert interviews and multidimensional unfolding can help identify important dimensions that are used as common judgment criteria for different stimuli. Even though multidimensional scaling has been used before in aesthetic research (e.g., Berlyne & Ogilvie, 1974; O'Hare, 1976) it has not been applied to the identification of relevant judgment dimensions of stimuli and their use for stimulus construction before. Now that it has been shown how existing methods can be used for stimulus and scale construction, aesthetic research should use the described approaches, especially when the use of systematically varied stimuli representing aesthetic objects is required. Sets of stimuli constructed in this way can be used in future research in various kinds of experimental settings to investigate the impact of other psychological (and non-psychological) factors on aesthetic judgments. One such use is demonstrated in constructing the VAESO. Specifically, in the present work, relevant aesthetic dimensions were identified for four specific object classes. The same approach could be used to investigate relevant aesthetic judgment dimensions for other everyday objects. This would allow constructing other systematically varied stimuli representing everyday objects of different object classes. Conducting research using all kinds of objects classes would not only allow generalizing the present results to other everyday objects. It would also allow investigating if aesthetic

judgments differ for different kinds of aesthetic objects. For example, one possible research aim could be to investigate if aesthetic judgments for objects belonging to object classes with more aesthetic qualities (such as paintings) differ from aesthetic judgments for objects belonging to object classes with more functional qualities (such as cutlery). Also, the methods used in the present research for constructing a scale for measuring aesthetic sensitivity for everyday objects can be used for stimulus and scale construction in other domains. Stimuli and scales for measuring aesthetic sensitivity for works of arts, for environmental stimuli such as rooms (as demonstrated for offices here) or every other class of objects that might be of interest could be constructed in the demonstrated way. Furthermore, using systematically varied stimuli not only allows assessing aesthetic sensitivity but also to investigate all kinds of aesthetic judgments. Furthermore, the used techniques are not limited to stimulus construction in aesthetic research but can be adapted to research in general that deals with the systematic variation of visual stimuli.

Possible critiques notwithstanding, the methods used in the present research seem very fruitful methods for future empirical aesthetic research. Multidimensional unfolding can help identifying important dimensions of the aesthetic judgment. Conjoint analysis can be used to construct new stimuli and to investigate the relative importance of aesthetic dimensions. Moreover, conjoint analysis might also be used for further analysis as is widely done in consumer research. For instance, in addition to evaluating main effects and interaction effects between different attributes of a set of stimuli, part-worth utilities could be used in regression analyses to investigate the relationship between the importance of aesthetic properties of stimuli as evaluated by conjoint analysis and individual difference variables such as personality, expertise, and others.

Future research as it relates more closely to the research reported in the present work needs to investigate the reliability and validity of the VAESO. Instead of using rating scales for the evaluation of the aesthetic values of the stimuli, the stimuli should be rank ordered within each object class. The new external standard that was developed in the present research could then be used in the described way to evaluate an individual’s aesthetic sensitivity. Further research might also reduce the scale such that it can be applied easily in various research contexts.

As far as the new external standard developed in the present research is concerned, it needs more investigation before it can be considered a valid external standard. Future research

must for example show whether the rank orders identified for the aesthetic dimensions are consistent over different samples.

Another important investigation concerns the different kinds of judgments that can be considered aesthetic judgments. In the present research, different kinds of aesthetic judgments were used. For example, the UgBeaScale asks individuals for their *affective* response towards the given stimuli. In contrast, in the research described in Chapters 2 to 4 individuals are asked for a rather *cognitive* response. In Chapter 2, aesthetic *preference* judgments were used to identify relevant aesthetic dimensions. In the research described in Chapter 3 and 4 participants were asked not to state their personal preference but how beautiful an object is in an "objective" sense. In the present research, these different judgments were evaluated in separate studies and the type of judgment was used that appeared most applicable to the context and respective scale the respective judgment seemed to make the most sense. However, it seems important to investigate if and how affective and cognitive aesthetic judgments differ and how they are related to each other.

Relatedly, on a conceptual level the research on aesthetic sensitivity and more broadly empirical aesthetic research overall needs to be build on a theoretical framework. Recently, Leder and colleagues (Leder, Belke, Oeberst, & Augustin, in press) proposed a model that integrates various findings of aesthetic research and that integrates aspects of emotional and cognitive aesthetic responses. The model specifies two types of output of aesthetic processes, namely aesthetic emotion and aesthetic judgments. Aesthetic emotion might be described as feelings of pleasure or happiness (but might also be negative) that occur particularly in the evaluation stage. This kind of aesthetic response might be what is assessed when participants are asked for their affective response towards given stimuli as in the research described in Chapter 1 of the present thesis. Following the model, aesthetic judgments refer to the object-related cognitive part of aesthetic processing. This kind of aesthetic response is assessed in the research described in Chapters 2 to 4 when participants are asked for their preference judgments or to evaluate the aesthetic value of objects in a rather objective way. Even though the model primarily focuses on art-specific aesthetic experiences, it includes other objects as aesthetically relevant and is in this way a model concerning all kinds of visual aesthetics (see Leder et al., in press). The different scales developed in the present research might be seen as approaches to measure the aesthetic emotion (UgBeaScale) and the aesthetic judgment (VAESO) as outcomes of an aesthetic process. Future research must show if and how the research presented in this thesis can be integrated into the model by Leder and colleagues.

References

Aickin, M., & Gensler, H. (1996). Adjusting for multiple testing when reporting research results: The Bonferroni vs Holm Methods. *American Journal of Public Health, 86*(5), 726-728.

Arnheim, R. (1985). The other Gustav Fechner. In S. Koch & D. E. Leary (Eds.), *A century of psychology as science.* New York: McGraw-Hill.

Bamossy, G., Scammon, D. L., & Johnston, M. (1983). A Preliminary Investigation of the Reliability and Validity of an Aesthetic Judgment Test. *Advances in Consumer Research, 10,* 685-690.

Bennett, J. F., & Hays, W. L. (1960). Multidimensional unfolding: Determining the dimensionality of ranked preference data. *Psychometrika, 25*(1), 27-43.

Berlyne, D. E. (1963). Complexity and incongruity variables as determinants of exploratory choice and evaluative ratings. *Canadian Journal of Psychology, 17,* 274-290.

Berlyne, D. E. (1970). Novelty, complexity and hedonic value. *Perceptual and Psychophysics, 8,* 279-286.

Berlyne, D. E. (1971). *Aesthetics and Psychobiology*. New York: Meredith Corporation.

Berlyne, D. E. (1974a). *Studies in the New Experimental Aesthetics: Steps toward and objective Psychology of Aesthetic Appreciation*. Washington, D.C.: Hemisphere Publishing Corporation.

Berlyne, D. E. (1974b). Verbal and Exploratory Responses to Visual Patterns varying in Uncertainty and in Redundancy. In D. E. Berlyne (Ed.), *Studies in the New Experimental Aesthetics: Steps toward an objective Psychology of Aesthetic Appreciation* (pp. 121-158). New York: John Wiley & Sons.

Berlyne, D. E., & Ogilvie, J. C. (1974). Dimensions of perception of paintings. In D. E. Berlyne (Ed.), *Studies in the new experimental aesthetics.* (pp. 181-226). Washington: Hemisphere.

Berlyne, D. E., Robbins, M. C., & Rhompson, R. (1974). A cross-cultural study of exploratory and verbal responses to visual patterns varying in complexity. In D. E. Berlyne (Ed.), *Studies in the new experimental aesthetics: Steps toward an objective Psychology of Aesthetic Appreciation* (pp. 259-278). New York: John Wiley & Sons.

Birkhoff, G. D. (1933). *Aesthetic Measure.* Cambridge, Mass.: Harvard University Press.

Bloch, P. H. (1995). Seeking the ideal form: Product design and Consumer Response. *Journal of Marketing, 59*(3), 16-29.

Bloch, P. H., Brunel, F. F., & Arnold, T. J. (2003). Individual Differences in the Centrality of Visual Product Aesthetics: Concept and Measurement. *Journal of Consumer Research, 29*, 551-565.

Borg, I., & Groenen, P. (1997). *Modern Multidimensional Scaling: Theory and Practice.* New York: Springer-Verlag.

Boring, E. G. (1942). *Sensation and Perception in the History of Experimental Psychology.* New York: Appleton-Century-Crofts.

Bortz, J. (1978). Psychologische Ästhetikforschung: Bestandsaufnahme und Kritik. *Psychologische Beiträge, 20*, 481-508.

Bourdieu, P. (1979). *La Distinction. Critique sociale du jugement.* Paris: Les Editions de Minuit.

Brunel, F. F. (1998). *The Psychology of Product Aesthetic: Antecedents and Individual Differences in Product Evaluations.* University of Washington.

Cattell, R. B. (1966). The scree test for number of factors. *Multivariate Behavioral Research, 1*, 245-276.

Chan, J., Eysenck, H. J., & Götz, K. O. (1980). A new Visual Aesthetic Sensitivity Test: III. Cross-cultural comparison between Hong Kong Children and Adults, and English and Japanese samples. *Perceptual and Motor Skills, 50*, 1325-1326.

Child, I. L. (1962). Personal preferences as an expression of aesthetic sensitivity. *Journal of Personality, 30*, 496-512.

Child, I. L. (1964). Observations on the Meaning of some Measures of Esthetic Sensitivity. *Journal of Psychology, 57*, 49-64.

Child, I. L. (1965). Personality correlates of esthetic judgment in college students. *Journal of Personality, 33*, 476-510.

Childers, T. L., Houston, M. J., & Heckler, S. E. (1985). Measurement of Individual Differences in Visual versus Verbal Information Processing. *The Journal of Consumer Research, 12*(2), 125-134.

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis*: Hillsdale, NJ: Erlbaum.

Coombs, C. H. (1964). *A theory of data.* New York: John Wiley & Sons.

Crannell, C. W. (1953). The Validity of certain measures of Art Appreciation in Relation to a Drawing Task. *The Journal of Psychology, 35*, 131-142.

Cronbach, L. J. (1951). Coefficient alpha and the Internal Structure of Tests. *Psychometrika, 16*, 297-334.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*, 349-354.

Cupchik, G. C. (1986). A Decade After Berlyne. New Directions in Experimental Aesthetic. *Poetics, 15*, 345-369.

Cupchik, G. C. (1994). Emotion in aesthetics: Reactive and reflective models. *Poetics, 23*, 177-188.

Eisermann, J. (2002). *Strukturerkennende Verfahren bei Daten aus offenem Antwortformat. Ein Beitrag zur Psychologie des Wunsches.* Free University of Berlin. http://www.diss.fu-berlin.de/2004/97/.

Eysenck, H. J. (1940). The General Factor in Aesthetic Judgments. *British Journal of Psychology, 31*, 94-102.

Eysenck, H. J. (1941). 'Type'-Factors in Aesthetic Judgments. *British Journal of Psychology, 31*, 262-270.

Eysenck, H. J. (1947). *Dimensions of personality.* London: Routledge & Kegan Paul.

Eysenck, H. J. (1965). Aesthetics and Personality. *Exakte Ästhetik, 1*(35-50).

Eysenck, H. J. (1967). Factor-Analytic Study of the Maitland Graves Design Judgment Test. *Perceptual and Motor Skills, 24*, 73-74.

Eysenck, H. J. (1970). An Application of the Maitland Graves Design Judgment Test to Professional Artists. *Perceptual and Motor Skills, 30*, 589-590.

Eysenck, H. J. (1983). Visual Aesthetic Sensitivity and its Measurement. In M. Ross (Ed.), *The Arts: A way of knowing* (pp. 105-125). Oxford: Pergamon Press.

Eysenck, H. J. (1988). Personality and Scientific Aesthetics. In F. H. F. a. R. W. Neperud (Ed.), *The foundation of aesthetics, arts and art education* (pp. 117-160).

Eysenck, H. J., & Castle, M. (1970). Training in Art as a Factor in the Determination of Preference Judgments for Polygons. *British Journal of Psychology, 61*(1), 65-81.

Eysenck, H. J., & Castle, M. (1971). Comparative Study of Artists and Nonartists on the Maitland Graves Design Judgment Test. *Journal of Applied Psychology, 55*(4), 389-392.

Eysenck, H. J., Castle, M., Averill, J. R., Stanat, P., & More, T. A. (1970). A Factor-Analytic Study of the Barron-Welsh Art Scale. *The Psychological Record, 20*, 523-525.

Eysenck, H. J., Goetz, K. O., Long, H. Y., Nias, D. K. B., & Ross, M. (1984). A new Visual Aesthetic Sensitivity Test - IV. Cross-cultural comparisons between a Chinese sample from Singapore and an English sample. *Personality and Individual Differences, 5*(5), 599-600.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the Use of Exploratory Factor Analysis in Psychological Research. *Psychological Methods, 4*(3), 272-299.

Fechner, G. T. (1876). *Vorschule der Ästhetik*. Leipzig: Breitkopf & Härtel.

Floyd, F. J., & Widaman, K. F. (1995). Analysis in the Development and Refinement of Clinical Assessment Instruments. *Psychological Assessment, 7*(3), 286-299.

Frith, C. D., & Nias, D. K. B. (1974). What determines aesthetic preferences? *The Journal of General Psychology, 91*, 163-173.

Frois, J. P., & Eysenck, H. J. (1995). The Visual Aesthetic Sensitivity Test Applied to Portuguese Children and Fine Arts Students. *Creativity Research Journal, 8*(3), 277-284.

Götz, K. O., Borisy, A. R., Lynn, R., & Eysenck, H. J. (1979). A new Visual Aesthetic Sensitivity Test: I Construction and Psychometric Properties. *Perceptual and Motor Skills, 49*, 795-802.

Graves, M. (1948). *Design Judgment Test*. New York: Psychological Corp.

Green, P. E., & Rao, V. R. (1971). Conjoint Measurement for Quantifying Judgmental Data. *Journal of Marketing Research, 8*, 355-363.

Green, S. B., & Srinivasan, V. (1978). Conjoint Analysis in Consumer Research: Issues and Outlook. *Journal of Consumer Research, 5*(September), 103-121.

Green, S. B., & Srinivasan, V. (1990). Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice. *Journal of Marketing, 54*(4), 3-19.

Green, S. B., & Wind, Y. (1975). New Way to measure Consumer Judgments. *Harvard Business Review, 53*, 107-117.

Hair, J. F. J. (1998). Conjoint Analysis. In J. F. J. Hair (Ed.), *Multivariate data analysis*. New Jersey: Prentice-Hall Inc.

Hays, W. L., & Bennett, J. F. (1961). Multidimensional Unfolding: Determining configuration from complete rank order preference data. *Psychometrika, 26*(2), 221-238.

Heckler, S. E., Childers, T. L., & Houston, M. J. (1993). On the Construct Validity of the SOP Scale. *Journal of Mental Imagery, 17*(3 & 4), 119-132.

Holbrook, M. B., & Zirlin, R. B. (1985). Artistic Creation, Artworks, and Aesthetic Appreciation: Some Philosophical Contributions to Nonprofit Marketing. *Advances in Nonprofit Marketing, 1*, 1-54.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scand J Statistics, 6*, 65-70.

Iwawaki, S., Eysenck, H. J., & Götz, K. O. (1979). A new Visual Aesthetic Sensitivity Test (VAST): II. Cross-cultural comparison between England and Japan. *Perceptual and Motor Skills, 49*, 859-862.

Jöreskog, K. G., & Sörbom, D. (1999). *LISREL 8.3*. Chicago, IL: Scientific Software International, Inc.

Karpowicz Lazreg, C., & Mullet, E. (2001). Judging the pleasantness of form-color combinations. *American Journal of Psychology, 114*(4), 511-533.

Karwoski, T. F., & Christensen, E. O. (1926). A Test for Art Appreciation. *Journal of Educational Psychology, 17*, 187-194.

Kim, J.-O., & Mueller, C. W. (1978). *Factor Analysis: Statistical Methods and Practical Issues.* Newbury Park: Sage Publications.

Koffka, K. (1935). *Principles of Gestalt psychology.* New York: Harcourt, Brace.

Kohli, R. (1988). Assessing Attribute Significance in Conjoint Analysis: Nonparametric Tests and Empirical Validation. *Journal of Marketing Research, 25*, 123-133.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971a). Additive Conjoint Measurement. In *Foundations of Measurement* (Vol. 1, pp. 245-315). New York: Academic Press.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971b). *Foundations of Measurement* (Vol. 1, Additive and Polynomial Representations). New York: Academic Press.

Krosnick, J. A. (1999). Survey Research. *Annual Review of Psychology, 50*, 537-567.

Kruskal, J. B. (1964). Nonmetric multidimensional scaling: A numerical method. *Psychometrika, 29*(1), 1-27.

Kruskal, W. H., & Wallis, W. A. (1952). Use of Ranks on One Criterion Variance Analysis. *Journal of American Statistical Association, 47*, 583-621.

Leder, H., Belke, B., Oeberst, A., & Augustin, D. (in press). A model of aesthetic appreciation and aesthetic judgment. *British Journal of Psychology*.

Luce, R. D., & Tukey, J. W. (1964). Simultaneous Conjoint Measurement: A New Type of Fundamental Measurements. *Journal of Mathematical Psychology, 1*, 1-27.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample Size in Factor Analysis. *Psychological Methods, 4*(1), 84-99.

Martindale, C., Moore, K., & Borkum, J. (1990). Aesthetic preference: Anomalous findings for Berlyne's psychobiological theory. *American Journal of Psychology, 103*(1), 53-80.

McIver, J. P., & Carmines, E. G. (1981). Unfolding Theory. In *Unidimensional Scaling*. Newbury Parc, CA: Sage.

Meier, N. C. (1940). *The Meier Art Test, Art. Judgment*. Iowa City: Bureau of Educational Research and Service. Univ. of Iowa.

Metzger, W. (1953). *Gesetze des Sehens.* Frankfurt a.M.: von Waldemar Kramer.

O'Hare, D. (1976). Individual differences in perceived similarity and preference for visual art: A multidimensional scaling analysis. *Perception & Psychophysics, 20*(6), 445-452.

Ritterfeld, U. (1996). *Psychologie der Wohnästhetik: Wie es uns gefällt.* Weinheim: Psychologie Verlags Union.

Ritterfeld, U. (2002). Social Heuristics in Interior Design Preferences. *Journal of Environmental Psychology, 22,* 369-386.

Steiner, D. L. (1994). Figuring out factors: The use and misuse of factor analysis. *Canadian Journal of Psychiatry, 39,* 135-140.

Stöber, J. (1999). Die Soziale-Erwuenschtheits-Skala-17 (SES-17): Entwicklung und erste Befunde zu Reliabilität und Validität. *Diagnostika, 45*(4), 173-177.

Stöber, J. (2001). The Social Desirability Scale-17 (SDS-17): Convergent validity, discriminant validity, and relationship with age. *European Journal of Psychological Assessment, 17,* 222-232.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics.* (4th ed.). Boston: Allyn & Bacon.

Tobacyk, J., Bailey, L., & Myers, H. (1979). Preference for paintings and personality traits. *Psychological Reports, 45,* 787-793.

Veryzer, R. W. J. (1993). Aesthetic Response and the Influence of Design Principles on Product Preferences. *Advances in Consumer Research, 20,* 224-228.

Welsh, G. S. (1949). *Barron-Welsh Art Scale*. Palo Alto: Consulting Psychologists Press.

Welsh, G. S. (1987). *Manual of the Welsh Figure Preference Test*. Palo Alto, CA: Consulting Psychologists Press.

Wittink, D. R., Krishnamurthi, L., & Nutter, J. B. (1982). Comparing Derived Importance Weights Across Levels on Conjoint Results. *Journal of Consumer Research, 8*(March), 471-474.

Wittink, D. R., Krishnamurthi, L., & Reibstein, D. J. (1990). The Effect of Differences in the Number of Attribute Levels on Conjoint Results. *Marketing Letters, 1*(2), 113-129.