

A NEW VISUAL AESTHETIC SENSITIVITY TEST:
I. CONSTRUCTION AND PSYCHOMETRIC PROPERTIES

K. O. GÖTZ

*Academy of Fine Arts,
Düsseldorf*

R. LYNN

*University of Ulster
Coleraine*

A. R. BORISY

*Institute of Psychiatry
University of London*

H. J. EYSENCK

University of London

Summary.—This paper reports progress in construction of a new test of visual aesthetic sensitivity, differing from previous tests in that the items were drawn by an experienced artist of international reputation. There are 42 sets of 2 non-representational pictures, differing in that one of these has been changed by the incorporation of certain intentional design faults; there is thus a "right" and a "wrong" picture, and it is the task of the subject to discover the right answer. Eight practising artists validated the construction by agreeing 100% in their answers with the key. The test was administered to 111 male and female students, and 369 male and female children. There were no sex differences, no very significant correlations with intelligence, and very little by way of correlations with personality traits. The mean score of the children was 30, that of the students 35, a very significant difference. Within the group of children, however, there was no correlation with age. The distribution of scores was skewed, with easy items (high scores) predominating. The difficulty levels of the items were similar for adults and children, male and female. The internal reliability of the test was .84 for the adults. The retest reliability of the test was .70 for girls and .32 for boys, but the samples were small.

There are several tests purporting to measure aesthetic sensitivity in the visual field, e.g., the Maitland Graves Design Judgment Test; these have not on the whole been very successful. The Graves (1946) test has been analyzed in detail by Eysenck (1967, 1970), by Eysenck and Castle (1971), and by Götz and Götz (1974), who demonstrated errors in design and lack of validity in discriminating artists from non-artists. The Barron-Welsh Art Scale (Welsh & Barron, 1963; Welsh, 1975) has a rather different purpose, although it too shows lack of homogeneity (Eysenck & Castle, 1970a). The major drawback of these and other similar tests (such as the Meier-Seashore) is that the stimuli are clearly of low or no artistic interest; artists in our experience often refuse point-blank to even look at the drawings as they consider them completely lacking in aesthetic value. The test here introduced represents the combined efforts of a painter and experimental psychologists, and it is hoped that the special skills of the former will improve the quality of the stimuli, while those of the latter will safeguard against the constructional errors vitiating some of the previously available tests.

The test consists of 42 pairs of drawings, one of which is considered "right" and the other "wrong." The score on the test is the number of "right" answers given in response to the instructions, which are given below. The instructions emphasize selection in terms of better configuration or "Gestalt" rather than simple preference, as will be appreciated on reading the text. We have found that usually the two correlate quite highly, but occasionally people high on the personality trait of psychoticism (Eysenck & Eysenck, 1976) prefer the drawing they themselves consider less "good" in these terms.

Instructions are as follows.

You will be shown a series of 42 pairs of designs; each pair consists of two rather similar designs. One of these is constructed to be more *harmonious* in its configuration than the other. Look carefully at the two designs, and you will see that the less harmonious design contains errors and faults; this judgment is based on the unanimous judgment of a group of experts, painters, and graphic artists. Your task is to discover which of the two designs is the better one, i.e., the more harmonious. Sometimes the "better" design will be on the right, sometimes on the left. Look carefully, and take your time before coming to a decision. It is only when all subjects taking part in the experiment have made their decision and have written 'R' or 'L' after the corresponding number of the design, that we will go on to the next design. The particular number of each design being shown will be called out.

Note that you are *not* asked to say which design you find more *pleasant*. Your task is to discover which of the designs is the more *harmonious* one. If there are any questions, please ask them now.

The nature of the test can best be clarified by stating first of all the nature of the criterion, and then illustrating it by some examples from the test. The artist first drew the "good" picture and then made certain alterations which intentionally incorporated faults in the picture. Having completed the set of pairs, he then asked eight well-known painters to go through the test and make their choices. A pair was only accepted if all eight judges gave unanimous agreement. Thus we have a professional criterion as the basis of the scoring procedure. We shall see later to what extent this criterion agrees with non-artistic preference.

Three examples are given below from the test, illustrating easy, middling and difficult items (Fig. 1). The difficulty level of an item is judged in terms of the proportion of correct choices made by a random group of subjects, with 50% constituting the chance level (impossibly difficult) and 100% constituting the opposite extreme of impossibly easy. Item 27 is an example of an easy item (average proportion of correct answers 95%), Item 9 constitutes a difficult item (average proportion of correct answers 60%), and Item 3 constitutes a middling item (average proportion of correct answers 79%). The proportions given are for a sample of 111 British University students without any special artistic training.

The mean score for this group was 35.79 for the 38 males, and 34.68 for

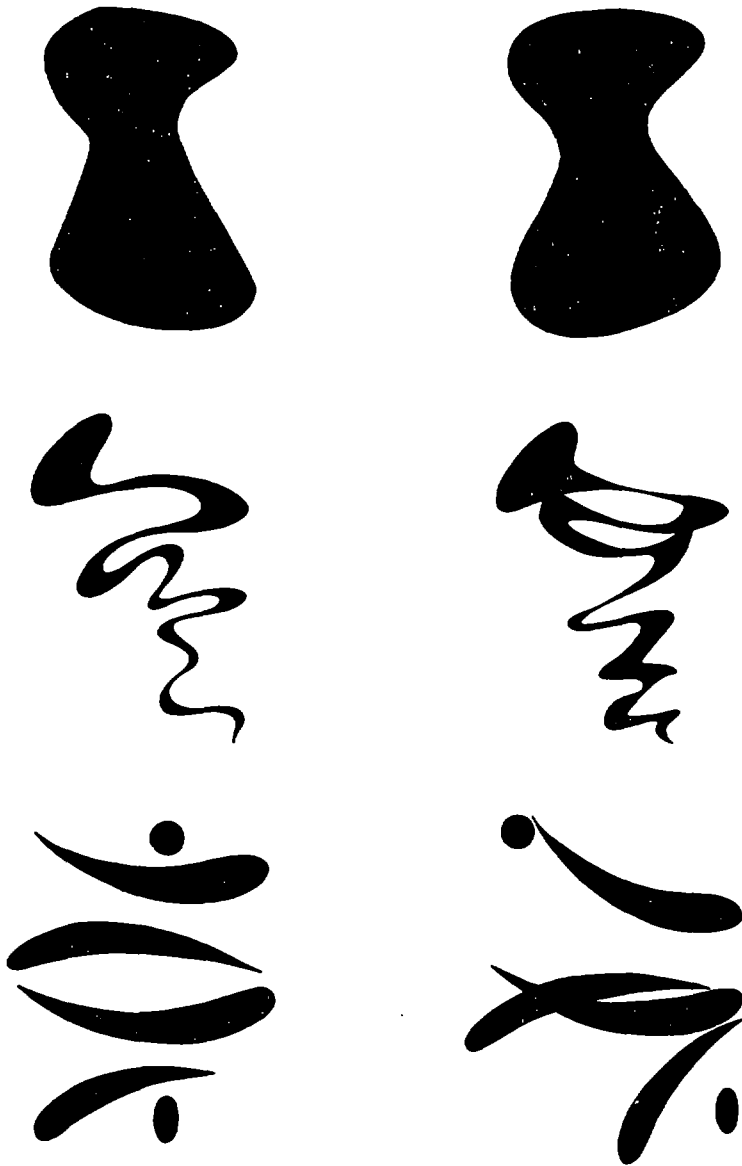


FIG. 1. Easy (No. 27, top), medium difficult (No. 3, middle), and difficult (No. 9, bottom) items of the Aesthetic Sensitivity Scale

the 73 females, with *SDs* of 4.66 and 5.26, respectively. Subjects were tested in small groups, with no time limit. It will be clear that the scores are highly skewed, with many high scores and few low ones. This feature is universal in

all the applications of the test so far made and may make necessary a change in the test structure, introducing more difficult items and withdrawing some of the easier items. However, as the test is intended for cultural comparison, among other things, and for the testing of children, it seemed wisest to err on the side of making the test too easy rather than too difficult. It seemed possible that a test appropriate for adult Western students might be too difficult for other groups.

The internal reliability of a test such as this is obviously of vital importance. Unless it is reasonably high no homogeneity of items can be assumed, and no identity of underlying trait hypothesized. The split-half reliability, corrected, was .84, which is quite high for a test of only 42 items. The average difference in score for split-halves was only 1 point. We can thus assume that whatever the test is measuring is some homogeneous quality which is being measured to approximately equal extent by arbitrarily selected subsections of the total test.

Scores for the adult population were correlated with scores on the Eysenck Personality Questionnaire (Eysenck & Eysenck, 1975), in order to see to what extent there might be a relationship between aesthetic sensitivity and personality. Table 1 shows the observed correlations, for males and females separately. It will be seen that only the negative correlation with Psychoticism has any claim to attention, and in view of the nature of this factor (Eysenck & Eysenck, 1976) the direction of this correlation (high Psychoticism scores—poor aesthetic sensitivity) is not surprising. The other correlations are essentially zero, indicating no relationship of any kind.

TABLE 1
PEARSONIAN CORRELATIONS BETWEEN AESTHETIC SENSITIVITY SCALE AND
PERSONALITY, FOR MALES AND FEMALES SEPARATELY

Measure	Males	Females
Psychoticism	-.21	-.37
Extraversion	.16	-.14
Neuroticism	.03	-.01
Lie Scale	.02	.00

We next turn to the scores of male and female children, tested in a variety of schools near Cambridge. There were 204 male children (32 aged 11 yr., 59 aged 12 yr., 65 aged 13 yr., 23 aged 14 yr., and 25 aged 15 yr.). There were also 165 female children (19 aged 11 yr., 54 aged 12 yr., 56 aged 13 yr., 13 aged 14 yr., and 23 aged 15 yr.). Fig. 2 gives the distribution of the combined scores for boys and girls; the skewing we have already drawn attention to in connection with the adult scores is quite obvious again. The mean total score is 30.4, with a median of 31, and a *SD* of 4.88. This is significantly lower than

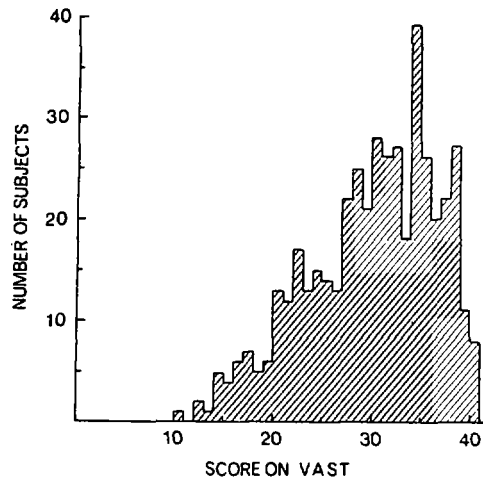


FIG. 2. Skewed distribution of scores on Aesthetic Sensitivity Test for sample of 454 boys and girls, aged 11-15 yr.

the score of the students, as might have been expected from immature children, but the difference is perhaps less impressive than one might have expected.

The children were also administered an intelligence test, and the Junior Eysenck Personality Questionnaire (Eysenck & Eysenck, 1975). All the correlations were insignificant, in spite of the quite large numbers involved, suggesting that intelligence and personality had little effect on the VAST score. Correlations with Psychoticism were negative for both boys ($-.08$) and girls ($-.04$) but clearly well below the values found for the adult groups. Correlations with IQ were both positive ($.07$ and $.20$), with the second value significant if taken in isolation. The possibility remains that intelligence may have a rather weak connection with aesthetic sensitivity.

It is interesting to see to what extent the position of the test items determines their evaluation, and accordingly 106 boys and girls were tested with the test slides in order. Under these conditions the score for this group was 31.92, just significantly above the mean for the normal procedure; however, as the sample tested differed of course from the sample tested with the orthodox version, it seems likely that in truth the positioning of the slides makes very little difference. Thirty-three boys and girls were tested with the test in the right position and then with the reversed test. The correlation was $.54$, and the scores were 34.03 and 34.94. This difference is not significant and suggests again that positioning makes very little if any difference. When the standard form was given on test and retest the means were 28.54 and 28.09, an insignificant fall.

Test-retest data are available on 47 girls and 42 boys, aged 12 to 14 yr. Correlations are $.70$ and $.32$, with the former quite satisfactory, and the latter too

low for practical purposes. Further retesting is obviously required to assess this discrepancy. Repetition of this test produces a good deal of boredom, more perhaps for the boys than for the girls. This factor may explain the disappointingly low retest reliability of the boys. The time elapsing between test and retest in all these studies was a matter of weeks rather than of months.

The difficulty levels of the items were calculated for all our groups and compared. As scores did not correlate significantly with age for either boys or girls, all the ages for the children were thrown together, but although there were also no sex differences in scores, we have kept the sexes separate. We thus have four groups: male children (A), female children (B), male adults (C), and female adults (D). The observed correlations between difficulty scores on the 42 items for these four groups are given in Table 2. They are all quite high and significant, showing that an item which was difficult for one group was also difficult for another.

TABLE 2
CORRELATIONS BETWEEN DIFFICULTY LEVELS OF ITEMS FOR FOUR POPULATIONS

Group	A	B	C	D
Male Children, A		.90	.70	.66
Female Children, B			.63	.71
Male Adults, C				.65
Female Adults, D				

The correlation between the two child groups is much the highest, perhaps because the groups in question were much bigger than the adult groups. We may perhaps regard this correlation as the most meaningful one in the table.

A factor analysis was carried out on the intercorrelations between items, for the boys and girls separately. Five factors with eigenvalues above 1 were extracted, but there was such a drop in the variance contributed by the second and subsequent factors that only the first factor has much meaning. Eigenvalues for the first factor, for boys and girls, respectively, were 5.37 and 5.42; the second factor had eigenvalues of 1.97 and 2.06, respectively. The nature of the first factor is identical for both sexes and very clear in interpretation. Difficult items have positive loadings, and easy items have negative loadings. Thus this is a "difficulty" factor, whose prominence indicates that apart from the general trait measured by the test there is little specific variance shared between items in this test. No interpretation proved possible of the second and subsequent factors, either unrotated or rotated.

DISCUSSION

The variable studied in this investigation, visual aesthetic sensitivity, constitutes a conceptual formulation deriving from a series of experiments carried out

many years ago (Eysenck, 1940, 1941a, 1942, 1965) regarding the nature of "taste" or good aesthetic judgment. This factor was found independent of another, related more to complexity of design, and similar to the concept studied in the Barron-Welsh Art Scale (Welsh, 1975). There is quite a literature on this "type" factor (Eysenck, 1941b, 1941c). Much of the work in this field has used simple polygonal figures (Eysenck, 1968, 1971, 1972a, 1972b, 1972c; Eysenck & Castle, 1970b). The present study is the first in which a set of stimuli has been used which could be called aesthetic in a manner which would be meaningful to the painter or sculptor, and the results are clearly not dissimilar, in that the choices made can be explained in terms of a single aesthetic ability, largely independent of intelligence or personality, possessed to different degree by different persons, and responding to different difficulty levels in the stimuli displayed. This ability seems to owe little to explicit teaching, in view of the fact that there was no correlation with age in the school children, and only a relatively small difference in scores between children and students. It is predicted that this ability, presumably having a genetic foundation in the structure of the nervous system, would transcend national and cultural boundaries, so that subjects of other races would show similar scores, and difficulty levels for item pairs, as the sample here tested. Future papers in this series will deal with such comparisons.

REFERENCES

- EYSENCK, H. J. The general factor in aesthetic judgments. *British Journal of Psychology*, 1940, 31, 94-102.
- EYSENCK, H. J. The empirical determination of an aesthetic formula. *Psychological Review*, 1941, 48, 83-92. (a)
- EYSENCK, H. J. Personality factors and preference judgments. *Nature*, 1941, 148, 346. (b)
- EYSENCK, H. J. "Type" factors in aesthetic judgments. *British Journal of Psychology*, 1941, 31, 262-270. (c)
- EYSENCK, H. J. The experimental study of the "good Gestalt"—a new approach. *Psychological Review*, 1942, 49, 344-364.
- EYSENCK, H. J. Aesthetics and personality. *Exakte Ästhetik*, 1965, 1, 35-49.
- EYSENCK, H. J. Factor analytic study of the Maitland Graves Design Judgment Test. *Perceptual and Motor Skills*, 1967, 24, 73-74.
- EYSENCK, H. J. An experimental study of aesthetic preference for polygonal figures. *Journal of General Psychology*, 1968, 79, 3-17.
- EYSENCK, H. J. Application of the Maitland Graves Design Judgment Test to professional artists. *Perceptual and Motor Skills*, 1970, 30, 589-590.
- EYSENCK, H. J. Factors determining aesthetic preferences for geometrical designs and devices. *British Journal of Aesthetics*, 1971, 2, 154-166.
- EYSENCK, H. J. The development of aesthetic sensitivity in children. *Journal of Childhood Psychology and Psychiatry*, 1972, 13, 1-10. (a)
- EYSENCK, H. J. Personal preferences, aesthetic sensitivity and personality in trained and untrained subjects. *Journal of Personality*, 1972, 40, 544-557. (b)
- EYSENCK, H. J. Preference judgments for polygons, designs and drawings. *Perceptual and Motor Skills*, 1972, 34, 396-398. (c)
- EYSENCK, H. J., & CASTLE, M. A factor analytic study of the Barron-Welsh Art Scale. *Psychological Record*, 1970, 20, 523-525. (a)

- EYSENCK, H. J., & CASTLE, M. Training in art as a factor in the determination of preference judgments for polygons. *British Journal of Psychology*, 1970, 61, 65-81. (b)
- EYSENCK, H. J., & CASTLE, M. A comparative study of artists and non-artists on the Maitland Graves Design Judgment Test. *Journal of Applied Psychology*, 1971, 55, 389-392.
- EYSENCK, H. J., & EYSENCK, S. B. G. *Manual of the Eysenck Personality Questionnaire*. London: Hodder & Stoughton, 1975.
- EYSENCK, H. J., & EYSENCK, S. B. G. *Psychoticism as a dimension of personality*. London: Hodder & Stoughton, 1976.
- GÖTZ, K. O., & GÖTZ, K. The Maitland Graves Design Judgment Test judged by 22 experts. *Perceptual and Motor Skills*, 1974, 39, 261-262.
- GRAVES, M. *Design Judgment Test*. New York: Psychological Corp., 1946.
- WELSH, G. S. *Creativity and intelligence: a personality approach*. Chapel Hill: Institute for Research in Social Science, 1975.
- WELSH, G. S., & BARRON, F. *Barron-Welsh Art Scale*. San Diego, CA: Consulting Psychologists Press, 1963.

Accepted August 23, 1979.