J. Child Psychol. Psychiat., Vol. 13, 1972, pp. 1 to 10. Pergamon Press. Printed in Great Britain.

THE DEVELOPMENT OF AESTHETIC SENSITIVITY IN CHILDREN

H. J. EYSENCK
Institute of Psychiatry, University of London

INTRODUCTION

LITTLE research appears to have been done on the development of aesthetic sensitivity in children, using objective measures; this is no doubt due to the absence of agreement of the value of such measures, or the criteria to be adopted. The work of Child (1964) serves as an excellent example of the difficulties which arise in constructing such a test, and the doubts which attend its validation. He used pairs of paintings where one or the other had been judged superior by expert judges; thus the validity of the test depended entirely on the subjective judgements of these judges. In view of the changing standards applied to works of art, it is doubtful if such a criterion is acceptable from the psychological point of view, and Child's discovery that children tended significantly to prefer the worse rather than the better of the two pictures presented must make one doubt if these expert standards had in fact much relevance to their judgements. Preference judgments for paintings are obviously multidimensional; pictorial content, colour and other factors influence different people's preferences differently, and confusion is likely to attend efforts to construct an objective standard in this fashion. Our own approach has been rather different, and follows the general considerations presented by Eysenck (1957).

Our first intention had been to use in our studies the Maitland Graves (1946) Design Judgement Test, which consists of 90 sets of 2 or 3 simple non-representational designs, one of which in each case has to be preferred to the other(s). The standard from which the scoring key was derived was the unanimous agreement of groups of art teachers, and Graves showed that art students had much higher scores on the test than did non-art students; these differences were offered by him as evidence of validity. There are two reasons against using the test in the manner intended. In the first place, it was found that for young children 90 sets of designs are too many; the children get bored and cease to attend. Forty designs was found to be the largest number which could reasonably be used. Consequently we restricted ourselves to this number.

In the second place, it was discovered in preliminary studies with adult *S*s that Graves's contentions were not in fact borne out when art and non-art students were tested (Eysenck, 1967, 1970; Eysenck and Castle, 1971). Differences between the two groups were small or non-existent; they certainly did not even approximate to those given in his *Manual*. Even worse, the underlying belief of Graves that "the test measures certain components of aptitude for the appreciation or production of art structure by evaluating the degree to which a subject perceives and responds to the basic principles of aesthetic order—unity, dominance, variety, balance, con-

Accepted manuscript received 19 *June* 1971

tinuity, symmetry, proportion, and rhythm" was somewhat undermined when it was found in correlational studies of the 90 items concerned that there was no single, underlying general factor, but rather 4 or 5 quite independent factors; the fact that there are several independent factors makes it difficult to justify the derivation of a single score by simple summation over all 90 items. Fortunately one of the factors appeared to present a reasonably unitary set of designs, linked together by one single underlying principle; this factor also succeeded in producing a highly significant differentiation between art and non-art students. (The other factors failed in producing such a differentiation, thus making the test as a whole a worse selection instrument than the sub-set of items making up factor 1.) It was decided, accrdingly, to select our 40 items by reference to these factorial studies, and then to carry out further factorial studies on selected groups of art and non-art students, as well as on the children themselves, in order to arrive at a defensible set of items which could be said to give us a proper measure of aesthetic sensitivity.

As an additional item, we included a personality inventory in the study, in the hope that some information would be obtained on the relation between aesthetic sensitivity and personality. There is little information in the literature to guide us in the formulation of hypotheses, and consequently none are proposed here; our hope is that the data themselves will suggest such hypotheses. Also included in the study were ratings of the amount of instruction in visual art which the children had received in each case; as these ratings failed to correlate with any of our data they will not be referred to again in this paper.

EXPERIMENT

Subjects

Our *S*s were school children, mostly from small towns in the West Country, ranging from 7 to 15 yr; there were too few at each end to make results very meaningful for these ages, but for the sake of completeness their results are included. For the most part, however, we shall concentrate on ages 8–14 only. Boys and girls took part in almost equal proportions. The actual numbers involved are given below. Most of the children came from comprehensive schools. An attempt was made to relate scores to type of school, but without success. Our *S*s are obviously not in any sense a random sample of British school children, but there is no obvious principle of selection which would make them unrepresentative. Technically all were volunteers, but in effect whole classes were tested, with no-one opting out. For the sake of comparison, and in order to make a proper selection of items for scoring, 200 male and 200 female art students, and 200 male and 200 female non-art students were tested. These *S*s were selected with reference to receiving or not receiving professional training in the visual arts, and are similar to those described in a previous publication (Eysenck and Castle, 1971). All groups were similar in age, averaging 21 yr.

Method

Forty items of the Design Judgement Test were reproduced on slides, and projected to groups of children in a darkened room; they were instructed to write down on score sheets supplied to them which of the alternatives they preferred. They

were told that there were no right or wrong answers, and that we were interested in their preferences. The actual items used are identified below. Either preceding or following the presentation of the items the children filled in a personality inventory giving scores on *E* (extraversion), *N* (neuroticism), *P* (psychoticism), and *L* (a Lie scale). The *E* and *N* scales are very similar to those making up the Junior EPI (Eysenck, 1965). The construction of the *P* and *L* scales has been discussed elsewhere (Eysenck and Eysenck, 1969, 1970, 1971; Eysenck, Nias and Eysenck, 1971). The scale used has not yet been published, but closely resembles in its make-up the scales discussed in the above references.

RESULTS

Adults

It was hoped to use results of the analysis of adult data for the selection of suitable items for our children's scale. The following conditions were laid down for selection: (1) an item would only be included if the majority of art students preferred the "correct" design, as indicated in Graves's key, (2) an item would only be included if the majority of non-art students preferred the "correct" design, (3) an item would only be included if the "correct" design was preferred more frequently by art students than by non-art students, (4) an item would be included only if in the factor analysis carried out separately for art students and non-art students it had a loading of 0·30 or above on the first principal component in both analyses. Twenty-two items were found which fulfilled these conditions, with only minor departures. Thus 5 items just failed to show majority vote for the "correct" design among the non-art students; on 3 items the non-art students showed very slightly better agreement with the "correct" choice than the art group; 3 items had loadings of just below 0·30 for one of the two groups, but made up for this by having a loading of above 0·30 for the other group. It was decided in each case to retain the item because although strictly speaking it should have been excluded, nevertheless on the other criteria it was clearly a desirable and useful item. In none of the cases considered was the failure to reach the criterion even remotely statistically significant. Our analysis thus furnishes us with a scale of 22 items which may be used as a measure of aesthetic sensitivity; it is this scale which will be used in our analysis of the results obtained with children.

A little more information requires to be given regarding the results of our study of adult choices. Using product-moment correlations to obtain our matrices for the 400 males and the 400 female *S*s, we then extracted two factors by means of principal components methods. The first of these was identified as a general factor of aesthetic sensitivity, and the loadings of the 22 items finally chosen are shown in Table 1, as are the Eigenvalues of the factors in question. (A varimax rotation was carried out, but produced so little change in loading pattern that it was disregarded.) Table 1 also gives the numbers of the 22 items chosen, so that they can be identified in the Design Judgement Test. Items 4, 10, and 16 are given out of order at the end of this table because these items had loadings on one or other analysis of less than 0·30.

Scales made up in the manner described would be expected to be reasonably reliable measuring instruments, and accordingly alpha coefficients were calculated for our groups. For the males, reliabilities were 0·88 for the non-art and 0·80 for the

TABLE 1. EIGENVALUE AND FACTOR LOADINGS OF CHILDREN AND STUDENTS (C = CONTROLS, A = ARTISTS) ON SELECTED ITEMS OF THE MAITLAND GRAVES DESIGN JUDGEMENT TEST

	Male				Female					
Ages	9	10	12	13	9	10	12	13	C	A
Eigenvalues	4·55	4·15	4·98	6·49	4·72	5·39	6·55	6·12	7·63	5·74
Design No.										
6	0·35	0·21	0·46	0·30	0·46	0·31	0·67	0·59	0·56	0·40
8	0·26	0·21	0·50	0·42	0·49	0·33	0·37	0·46	0·52	0·38
21	0·49	0·66	0·67	0·53	0·48	0·35	0·53	0·56	0·58	0·53
25	0·34	0·26	0·38	0·48	0·35	0·27	0·07	0·58	0·32	0·43
28	0·34	0·38	0·21	0·24	0·19	0·37	0·25	0·56	0·48	0·36
29	0·39	0·54	0·52	0·60	0·45	0·52	0·61	0·50	0·62	0·53
42	0·36	0·41	0·33	0·40	0·42	0·46	0·32	0·45	0·42	0·31
44	0·64	0·38	0·53	0·63	0·36	0·39	0·28	0·54	0·51	0·33
47	0·25	0·40	0·41	0·56	0·48	0·44	0·37	0·51	0·56	0·56
48	1·51	0·56	0·60	0·64	0·60	0·65	0·50	0·54	0·66	0·56
50	0·41	0·39	0·50	0·58	0·41	0·40	0·63	0·29	0·60	0·46
51	0·39	0·30	0·25	0·45	0·32	0·48	0·46	0·31	0·61	0·38
53	0·32	0·11	0·44	0·32	0·30	0·37	0·47	0·18	0·52	0·36
57	0·53	0·29	0·52	0·63	0·52	0·53	0·61	0·55	0·65	0·66
59	0·64	0·33	0·53	0·68	0·45	0·58	0·65	0·56	0·67	0·58
61	0·64	0·60	0·41	0·58	0·52	0·58	0·47	0·49	0·69	0·54
66	0·59	0·50	0·34	0·57	0·54	0·52	0·40	0·63	0·54	0·52
69	0·46	0·25	0·42	0·57	0·49	0·44	0·60	0·29	0·55	0·40
72	0·12	0·19	0·20	0·40	− 0·19	0·36	0·43	0·47	0·46	0·40
4	0·27	0·43	0·21	0·26	0·39	0·38	0·26	0·05	0·29	0·32
10	0·18	0·18	0·19	0·45	0·12	0·36	0·47	0·30	0·48	0·29
16	0·04	0·03	0·35	0·45	0·21	0·29	0·36	0·39	0·42	0·26
P	0·12	0·07	− 0·18	0·00	− 0·23	− 0·43	− 0·25	− 0·12		
E	0·17	0·09	0·04	− 0·12	0·13	− 0·14	− 0·02	− 0·05		
N	− 0·02	0·02	− 0·24	− 0·08	0·14	− 0·11	− 0·20	0·01		
L	− 0·05	− 0·02	0·19	0·10	0·17	− 0·07	0·18	0·13		

art groups; for the females, reliabilities were 0·88 and 0·83. These figures are sufficiently high to give one confidence in the reliability of the measuring instrument. Each *S* was give a score on the basis of giving 1 point for each correct answer; mean scores were 17 and 19 for male non-art and art students respectively, and 18 and 19 for the female non-art and art students. (Exact means and *S.D.*s are given in Table 2.)

Children

Main results of this study are given in Table 2, which shows the number of children in each age and sex group, the reliability of the score (alpha coefficients), and the mean scores and *S.D.*s; for the sake of comparison, Table 2 also shows the results for the two adult groups. The letter *C* (controls) refers to the non-art group, the letter *A* to the art group. The number of *S*s is given in connection with each group; as mentioned before, there are too few children in the 7 and 15 yr groups to make the results meaningful, but they have been included for the sake of completeness. Scores do not attain reasonable reliability until the age of 9; they do not seem to change much after that, at least not in any consistent fashion. Boys and girls do not seem to differ with respect to reliability. The figures suggest that whatever the test is measuring in adults, it is measuring with only slightly less reliability in children of 9 yr and older.

TABLE 2. RELIABILITIES (ALPHA COEFFICIENTS), SCORES AND *S.D.*'S OF DIFFERENT GROUPS OF CHILDREN. ALSO GIVEN ARE NUMBERS OF CHILDREN IN EACH GROUP

Age	*N*	*r*	Score	*S.D.*
		Male		
7	11	0·27	16·36	1·38
8	41	0·54	16·57	1·65
9	78	0·75	16·13	2·14
10	77	0·71	15·95	2·00
11	46	0·58	15·91	1·70
12	97	0·78	16·64	2·22
13	112	0·86	16·68	2·66
14	35	0·76	16·76	2·19
15	30	0·76	16·57	2·18
C	200	0·88	17·06	2·89
A	200	0·80	18·87	2·15
		Female		
7	16	0·05	16·38	1·20
8	54	0·38	16·13	1·47
9	82	0·75	16·02	2
10	85	0·80	15·86	2·32
11	49	0·62	16·95	1·81
12	103	0·83	16·86	2·42
13	97	0·83	17·07	2·49
14	51	0·57	16·65	1·69
15	7	0·14	16·64	1·07
C	200	0·88	17·89	2·75
A	206	0·83	18·66	2·31

Figure 1 graphs the scores of children and adults, for both sexes combined. There appears to be a none-too-regular progression of scores with age, with all secondary school groups having higher scores than all primary school groups. Analysis of variance (excluding the 7 and 15 yr groups because of their small size) shows that sex is insignificant, both as a main variable and in interaction with age. Age is significant beyond the 0·001 level, and many individual comparisons are significant by *t*-test, at levels going up to $p < 0{\cdot}001$. Ages 9 and 10, compared with ages 12, 13 and 14 give the biggest p values. These results leave little doubt that aesthetic sensitivity, as measured here, increases with age and that this progression is not confined to one sex.

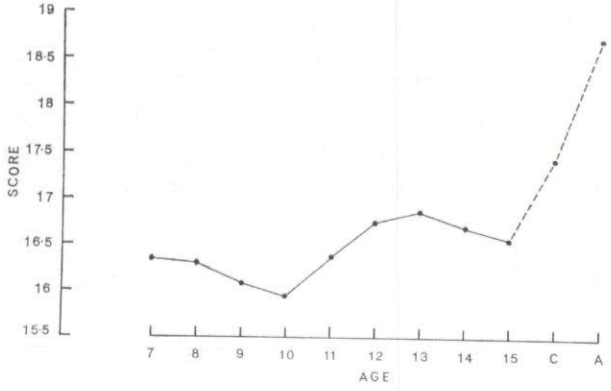

FIG. 1. Scores of children, non-art students and art students on 22-item scale of aesthetic sensitivity

It may be useful to add to the evidence from reliability further proof that the test measures much the same sort of thing in children as it measures in adults. Factor analyses were carried out on the 9-, 10-, 12- and 13-yr-old boys, and the 9-, 10-, 12- and 13-yr-old girls, exactly along the lines of the analyses already reported for the adults. (The other age groups had too few children in them to make factor analysis possible.) Results are reported in Table 1. It will be seen that the loadings of the items are all positive (with a single, insignificant exception among the 9-yr-old girls), and roughly similar to, though smaller than, the loadings for adults. Size of loadings seems to increase with age, as one might have expected, and so does the size of the Eigenvalues characteristic of each principal component. These figures suggest that the test measures a similar or identical factor in children as in adults, but that whatever psychological trait gives rise to this factor is less clearly structured in the younger children than in the older ones, and in children as a whole than in adults. The conclusions agree quite well with those derived from our study of means and reliabilities.

For our children groups, scores on the *P*, *E*, *N* and *L* variables of the questionnaire were included in the factor analysis, and the loadings of these scales on the "aesthetic sensitivity" factor are shown in Table 1. If we may treat these loadings as correlations, then several of them are significant at the 0·05 and 0·01 levels, and in addition there seems to be some consistency, even where correlations fall short of significance.

Correlations with *P* are almost universally negative, and frequently significantly so; we may conclude tentatively that the cold, impersonal, aggressive, unemotional individual that emerges from the items of the inventory has little aesthetic sensitivity, a conclusion which may be in good agreement with popular stereotypes of the "artistic temperament". *N*, too, appears to have a negative relation to aesthetic sensitivity, although this relation is somewhat less strong than that of *P*; replication would be required before much attention can be given to this finding. *L* and *E* do not correlate significantly with sensitivity in any of the groups; it is perhaps suggestive that *L* tends to have several positive correlations with sensitivity in the older groups. In view of the dearth of previous information in this field, none of these figures can be taken too seriously, but the suggestion remains that personality may correlate to some extent with aesthetic sensitivity.

Total scale

It might be thought that our results are in part an artefact produced by selection of a special set of 22 items from our total of 40. It may therefore be useful to report briefly on the results achieved with the total scale. The Graves numbers of the 18 additional items are: 5, 12, 14, 15, 18, 20, 22, 23, 24, 27, 30, 32, 40, 55, 76, 78, 79, 80. All had positive loadings on the "aesthetic sensitivity" factor, but these were smaller than those of the items included in the short scale. On many of these items the average preference for the "correct" design was less than 50 per cent, and often the "correct" design was preferred more frequently by the non-art students than by the art students. Taking the whole set of 40 items, there were still significant differences

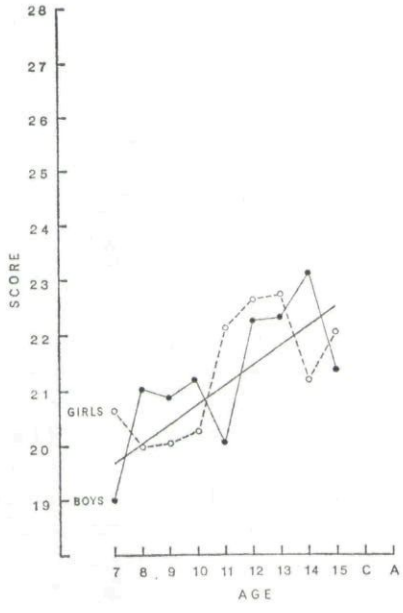

FIG. 2. Scores of male and female children on 40-item scale of aesthetic sensitivity. Regression line fitted by eye.

between the art students and the non-art students; with respect to the children, there was again a moderately regular improvement in scores with increasing age, regardless of sex. Figure 2 plots mean scores, for the children and adults. Analysis of variance again showed the progression of score with age to be significant. There is thus no reason to suspect that the selection of 22 items from the total scale of 40 used in the experiment had altered the main conclusion to be derived from the study; it may be hoped that selection has somewhat improved the test and made the results clearer than they would otherwise have been. Furthermore, the shortened test is psychometrically and psychologically more defensible than the longer one. But essentially both versions give similar results.

There is, however, one point on which the results from the 40-item scale deviate from those of the 22-item scale. With the short scale, all the children's groups were well above the chance level of approximately 11; with the long scale, the 7-yr-olds are just about at the chance level, and while the older groups exceed it, they do not do so by very much. This means that on the added items the children are more frequently "wrong" than "right", and this is indeed so; we have already mentioned that these added items also have much lower factor saturations, i.e. they correlate much less with the central items of the short scale. The added items are thus less "relevant" (they do not conform very well to our criteria), but they produce a more marked increase in score with age. It may be possible that these added items are more closely determined by formal teaching, but of course our data are not capable of providing us with evidence on this point.

It is possible to calculate a "difficulty index" for each item, for each group, by simply obtaining the proportion of "correct" answers for that item, in that group. The question arises whether these difficult indices are similar from group to group (as they would undoubtedly be in the case of I.Q. items), or whether there are marked differences between groups. Correlations were calculated between all 22 groups, and the results indicate some rather curious clusterings. Taking the adults first, we find that the male and female controls correlate highly together (0·79), as do the male and female art students (0·85). However the correlations between the art and non-art student groups are quite low. For the males, $r = 0{\cdot}07$, for the females $r = 0{\cdot}51$, and for the male–female comparisons the correlations are 0·28 and 0·28. Thus being or not being an art student makes a profound difference to the "difficulty level" of the items. This suggests that by using this information and weighting items differentially it would theoretically be possible to obtain a much better differentiation between the groups than was obtained by straightforward summing of scores. As there seems little point in making such a differentiation in practice, this line of argument has not been pursued. Instead, it seemed worth while to look at the correlations between the children's indices and those of the adults, with a view to seeing whether agreement would be higher with the art or non-art groups. The answer, in brief, seems to be that correlations throughout are low with both groups; children have difficulty indices almost entirely unrelated to those of adults. Some of the correlations are indeed negative, and none exceed 0·30. However, correlations between the different age groups within the children's sample are consistently high, rising to the 0·80s and 0·90s. There are thus three distinct groups which emerge from our analysis of difficulty indices: art students, non-art students, and children.

Sex does not cause any noticeable differentiation in this connection. The meaning of these groupings is not easy to understand, but the facts are so clear that it seemed desirable to state them briefly, in the hope that further work might throw some light on the problem.

DISCUSSION

The general tenor of the results seems to be in good accord with what commonsense might have predicted. If our test measures some component of aesthetic sensitivity, then art students are superior to non-art students; as far as children are concerned, sensitivity seems to improve as they grow older. The improvement is rather slight, even though statistically significant, and it is by no means as clear-cut as one might desire. Furthermore, the improvement seems independent of formal teaching; correlations between formal teaching and scores on the test were uniformly insignificant. Taking all these facts together, one might come to the conclusion that maturation was probably more influential than teaching in producing "correct" judgements on the test; indeed, the fact that even at the age of 7 children already have scores of over 16, when 11 represents the chance level, suggests that formal teaching cannot have had much influence on their ability. With a total range of 11 points (from chance level to perfect score), our 7-yr-olds have already come up almost to the half-way point on our scale, in the absence of any art teaching at all; this fact too suggests the presence of some genetic factors. Direct work with twins would seem indicated to verify this suspicion. The fact that differences between art and non-art students are quite small (just over one point on an 11-point scale) also suggests that formal teaching does not play much part in the genesis of the aesthetic response—quite apart from the fact that such differences as are found might have been present before students took up their courses, and may have been instrumental in determining choice of career, and likelihood of acceptance in art school.

Critics may raise doubts about the adequacy of the test. Criteria in this field are notoriously difficult to agree on, but we have made a special effort to construct a test which would meet several different criteria simultaneously. Determination of the "correct" answer for each item was made dependent on a number of criteria; (1) the answer was agreed on by all the experts in Graves's original group of judges; (2) the answer was agreed on by a majority of the art students in our sample; (3) the answer was agreed on by a majority of the non-art students in our sample; (4) the correct answer was given more frequently by art students; (5) the item correlated positively and highly with all the other items. This multiplicity of criteria, plus the fact that the scale has high internal reliability, does not guarantee that we have achieved a pure measure of aesthetic sensitivity, but it makes it more likely than a less stringent system of selection that our data will prove capable of replication, and that our conclusions will make psychological sense. In a rather novel field, with little previous research to guide one, more definitive conclusions are obviously unlikely to be forthcoming; it is to be hoped that the topic will be taken up by other workers in the educational field.

SUMMARY

Five hundred and twenty-seven male and 544 female children, aged 7–15, and

400 male and 400 female art and non-art students were administered a shortened version of the Maitland Graves Design Judgement test. Of the 40 items used, a scale of 22 items was selected on the basis of factor analysis, and a comparison of the scores of the art and non-art groups; this scale had high internal reliability, both for the students and the children. The scores of the children on this scale increased with increasing age, but the increase was small, although statistically very significant, and irregular; non-art students scored higher than the children, and art students higher than non-art students. Personality questionnaire scores were found to correlate with aesthetic sensitivity in the children. No correlations were found between sensitivity and amount of formal art teaching received at school. This, together with the relatively high scores achieved by quite young children of 7 and 8, suggests that sensitivity is not due to formal training in visual arts.

Acknowledgements—I am indebted to the S.S.R.C. for a grant which made this research possible, and to Miss Maureen Castle for help in the collection of data.

REFERENCES

CHILD, I. L. (1964) Developmental sensitivity to esthetic values. Yale University: Cooperative Res Program of the Office of Education U.S. Dept. of Health, Education and Welfare, No. 1748.

EYSENCK, H. J. (1957) The psychology of aesthetics. In *Sense and Nonsense in Psychology*, pp.308–339. Penguin Books, London.

EYSENCK, H. J. (1967) Factor analytic study of the Maitland Graves Design Judgement Test. *Percep. Motor Skills* **24**, 73–74.

EYSENCK, H. J. (1970) An application of the Maitland Graves Design Judgement test to professional artists. *Percep. Motor Skills* **30**, 589–590.

EYSENCK, H. J. and CASTLE, M. (1971) A comparative study of artists and non-artists on the Maitland Graves Design Judgement test. *J. Appl. Psychol.*, in press.

EYSENCK, H. J. and EYSENCK, S. B. G. (1970) A factor analytic study of the Lie Scale of the J.E.P.I. *Personal.* **1**, 3–10.

EYSENCK, H. J. and EYSENCK, S. B. G. (1971) Personality measurement in children: a dimensional approach. *J. Special Educ.* **4**, 261–268.

EYSENCK, S. B. G. (1965) *Manual of the Eysenck Personality Inventory*. University of London Press, London.

EYSENCK, S. B. G. and EYSENCK, H. J. (1969) "Psychoticism" in children. *Res. Educ.* **1**, 21–37.

EYSENCK, S. B. G., NIAS, D. K. B. and EYSENCK, H. J. (1971) On the interpretation of Lie Scale scores in children. *Br. J. Soc. Clin. Psychol.*, in press.

GRAVES, M. (1946) *Design Judgement Test*. Psychology Corporation, New York.