

Imageability and concreteness

JOHN T. E. RICHARDSON

Brunel University, Uxbridge, Middlesex, England

A factorial experiment was carried out to test the hypothesis that imageability and concreteness are alternative measures of the same verbal attribute. There was a highly significant interaction between the two scales and the individual stimulus items employed. Examples are given of items differentiating between the scales. It was concluded that imageability and concreteness should be distinguished, both experimentally and theoretically.

There now exists a considerable amount of evidence to support the contention that mental imagery is an important factor in many psychological tasks. This research has employed two different concepts, imageability and concreteness, as alternative measures of the image-arousing capacity of verbal material (e.g., Paivio, 1969, 1971b, pp. 77-80). These measures are derived from experimental situations where large numbers of subjects are asked to evaluate items on each of the two dimensions, rating them along seven-point scales (Paivio, Yuille, & Madigan, 1968). The instructions define an item's imageability as the extent to which it evokes a mental image, and they define its concreteness as the extent to which it can be experienced by the senses. These definitions are very similar, and it is not surprising that the two rated variables correlate highly with each other. Indeed, Paivio (1969, 1971b, p. 79) has maintained that they represent alternative measures of the same verbal attribute. However, this hypothesis is not to be tested by considering the correlation between the two variables, but by contrasting them within an experimental design.

A recent series of experiments (Richardson, 1975a, 1975c) has investigated the alternative hypothesis that imageability and concreteness represent distinct, though correlated, attributes. When the two variables were manipulated independently in immediate free recall, it was found that imageability had a positive effect upon the recall of items of low concreteness, but no effect upon the recall of items of high concreteness. In a subsequent experiment, it was found that imageability would affect

This research was supported by the Science Research Council and by the Medical Research Council of the United Kingdom. The experiment was carried out while the author was at the Laboratory of Experimental Psychology, University of Sussex. Professor A. Paivio kindly provided the list of nominalizations from which the material was selected, and E. C. C. Wynter, Headmaster of Haywards Heath Grammar School, gave permission for the questionnaires to be administered to his pupils. R. Q. Goodwin assisted in the selection of the material. Finally, I am grateful for the comments of W. F. Battig, who sponsors this paper and takes full editorial responsibility for its content. Requests for reprints should be sent to J. T. E. Richardson, Department of Psychology, Brunel University, Uxbridge, Middlesex, England.

the recall of items of high concreteness if the subjects were told to use imagery in their memorizing. This suggested that imagery was used to remember items of low concreteness but that, in the absence of specific instructions, it was not employed to remember items of high concreteness. Two further experiments studied the semantic categorization of verbal items and the latency of free associates to such items. In both studies, concreteness was a significant predictor of performance, but imageability was not. It was concluded that concreteness was a feature of lexical organization, and not a measure of the image-arousing quality of verbal material.

Nevertheless, without a demonstration of reliable differences over verbal items between imageability and concreteness, there is no assurance that these effects were not artifactual. (The converse also holds: Without experimental studies of the kind mentioned above, there would be no assurance that reliable differences between the two scales had any psychological importance.) The aim of this paper is to ask directly whether subjects evaluating an item's imageability produce different results from those evaluating that item's concreteness. It also considers the reliability of the two variables, and whether there are differences between males and females in their rating responses.

METHOD

Subjects

The subjects were 96 pupils at a grammar school; they were all aged between 16 and 18 years, and were in their penultimate year at school before entering university.

Materials

The pool of 360 English nominalizations (e.g., *migrating birds*, *painting gates*) employed by Paivio (1971a) was reduced to 244 items by excluding any phrase which, in the opinion of either the author or another judge, was clearly ambiguous or anomalous (Richardson, 1975b).

Procedure

The subjects were tested as a group. The phrases were presented in imageability and concreteness rating questionnaires of the form described by Paivio (1971a; Paivio et al., 1968). Four of the phrases were given as examples with the instructions, and the remaining 240 phrases were listed in a random fashion on 12 pages. Two groups of 12 male subjects and two groups of 12 female subjects completed each of the two kinds of

Table 1
Items Differentiating I and C

	I	C
Wandering thoughts	5.38	2.98
Thrilling sensations	5.35	3.10
Burning love	4.94	3.38
Needing love	5.21	3.71
Flirting glances	5.29	3.88
Frightening cliffs	5.44	4.08
Twinkling stars	6.02	4.75
Imagining pains	3.56	2.33
Flaring anger	4.77	3.56
Laughing faces	5.73	4.67
Building bridges	4.19	5.69
Stealing flags	2.63	4.29
Piercing ears	3.85	5.56
Minting coins	3.79	5.58
Demolishing villages	4.17	6.02
Trapping wolves	3.21	5.23
Publishing editors	2.27	4.46
Editing books	2.52	4.79
Making candles	3.52	5.81
Cashing bonds	2.81	5.19

questionnaire. The instructions were shown on the first page of the questionnaires, and the order of presentation of the other 12 pages to each of the 12 subjects in a group was determined by a Latin square.

RESULTS

The basic design of this experiment contained two random factors, subjects and items. The former was nested under two fixed factors, questionnaires and sexes, and one random factor, replications. Preliminary tests on the model for the analysis of variance showed that no term involving the factor of replications was significant ($F < 1$ in all cases), so each such term was pooled with its error term. There was a significant effect of questionnaires [quasi $F(1,116) = 4.43, p < .05$] and a significant effect of items [$F(243,22356) = 29.6, p < .001$]. There was a significant interaction between the effect of questionnaires and the effect of items [$F(243,22356) = 5.59, p < .001$], a significant interaction between the effect of sexes and the effect of items [$F(243,22356) = 1.60, p < .01$], and a significant interaction between the effect of questionnaires, the effect of sexes, and the effect of items [$F(243,22356) = 1.37, p < .01$]. Neither the main effect of sexes nor its interaction with questionnaires was significant (quasi $F < 1$ in both cases).

The mean rating of imageability (I) over 48 subjects and 244 items was 4.17; the mean rating of concreteness (C) was 4.47. The 10 items on which the males' I rating most exceeded the females' were: *ascending towers, awarding degrees, baying hounds, composing songs, disgusting ideas, inspiring structures, minting coins, overthrowing governments, tuning pianos, and trapping wolves*. The 10 items on which the females' I rating most exceeded the males' were: *crying babies, differing opinions, enlightening books, losing gloves, minding babies, ordering ambulances, persisting doubts,*

skipping meals, sleeping babies, and suppressing anger. The 10 items on which the males' C rating most exceeded the females' were: *astounding strengths, calculating results, detouring streams, disgusting ideas, existing situations, getting ideas, intriguing results, scaling cliffs, shirking obligations, and warring tribes*. The 10 items on which females' C rating most exceeded the males' were: *aggravating actions, complaining parents, dying monarchs, entering marriage, hearing rumors, lowing cows, laughing faces, reigning kings, thundering herds, and tuning pianos*. Table 1 shows the 10 items on which rated I and C diverged to the greatest extent in each direction, together with the means of the two variables on each item over 48 subjects.

The design of the experiment enables one to determine the reliability of I and C by computing the correlations over the 244 items between the two groups of 12 subjects of each sex who completed each questionnaire. On the imageability questionnaire, the males produced a correlation of +.81 and the females yielded +.83. On the concreteness questionnaire, the males produced a correlation of +.74 and the females yielded +.88. These are somewhat lower than the intergroup correlations of +.94 reported by Paivio et al. (1968). The correlations between the 24 subjects of each sex on each questionnaire were +.82 in the case of imageability and +.88 in the case of concreteness. The correlation between I and C derived from the 48 subjects who completed each questionnaire was +.68. Again, this is lower than the value of +.83 reported by Paivio et al. (1968) on the basis of ratings of 925 nouns.

DISCUSSION

Although rated imageability and concreteness correlate highly with each other, this study has shown a substantial difference between the two variables across different verbal items. Apart from an overall difference of .30 of a scale unit between ratings of I and C, certain items may be truly said to differentiate the two scales. In their own study, Paivio et al. (1968) claimed that the items for which I and C were most discrepant had common characteristics. They suggested that items for which I exceeded C tended to have strong emotional or evaluative connotations. On the other hand, they claimed that items for which C exceeded I were recognized by subjects as referring to concrete situations, but that these situations were experienced only infrequently. These suggestions apply reasonably well to the examples shown in Table 1. However, to attempt to characterize the discrepant items in this way is to recognize that the definitions of imageability and concreteness are logically distinct, and to acknowledge that the subjects' rating behavior may be reflecting this distinction. It is therefore inconsistent with the general theoretical attitude adopted by Paivio (1969) that the two variables are alternative measures of the same underlying process.

The results of the experiment described here also permit one to reject the idea that any discrepancies between I and C are caused by unreliable measurement. Not only is the experimental separation of the two variables not impractical (cf. Paivio, Yuille, & Smythe, 1966), but it is rendered legitimate by the highly significant interaction between the two scales and the different items employed, which shows that the discrepancies between I and C are statistically reliable. Further, this result supports a theoretical distinction between imageability and concreteness, and invites the exploration of this distinction

through experimental situations based upon the orthogonal manipulation of the two variables. The available evidence suggests that effects of imageability can be related to the mental process of imagery by their interaction with the effects of instructions to use imagery in psychological tasks, whereas concreteness appears to be an attribute of linguistic structure, insofar as it affects performance in experiments investigating semantic organization (Richardson, 1975a).

Within each of the scales, there appear to be reliable differences between males and females in their rating responses. Many of the examples illustrating this discrepancy can be easily understood in terms of differences in experience dictated by expectations, roles, and stereotypes prevalent in Western society. It would be surprising indeed if males and females did not differ in their ratings, and these effects should be remembered in future research. However, it is interesting that they are quantitatively small, and they are certainly much less important than the discrepancies between imageability and concreteness which have already been discussed.

The present study has shown that imageability and concreteness are highly reliable, that males and females agree well in their responses, and that there exists a high correlation between the two variables. However, the factorial design of the experiment has allowed a test of the hypothesis that they are alternative measures of the same attribute, and this hypothesis has been rejected. The two variables deviate significantly from each other, and items may be chosen which differentiate between them. It would therefore be fruitful to distinguish between imageability and concreteness in future research.

REFERENCES

- PAIVIO, A. Mental imagery in associative learning and memory. *Psychological Review*, 1969, **78**, 241-263.
- PAIVIO, A. Imagery and deep structure in the recall of English nominalizations. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 1-12. (a)
- PAIVIO, A. *Imagery and verbal processes*. New York: Holt, Rinehart & Winston, 1971. (b)
- PAIVIO, A., YUILLE, J. C., & MADIGAN, S. A. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph*, 1968, **76**(1, Pt. 2).
- PAIVIO, A., YUILLE, J. C., & SMYTHE, P. C. Stimulus and response abstractness, imagery, and meaningfulness, and reported mediators in paired-associate learning. *Canadian Journal of Psychology*, 1966, **20**, 362-377.
- RICHARDSON, J. T. E. Concreteness and imageability. *Quarterly Journal of Experimental Psychology*, 1975, **27**, 235-249. (a)
- RICHARDSON, J. T. E. Imagery and deep structure in the recall of English nominalizations. *British Journal of Psychology*, 1975, **66**, 333-339. (b)
- RICHARDSON, J. T. E. Imagery, concreteness, and lexical complexity. *Quarterly Journal of Experimental Psychology*, 1975, **27**, 211-223. (c)

(Received for publication January 22, 1976.)