STRATEGY TRAINING AND REASONING

Michael J. Lawson
The Flinders University of S.A.
and
John R. Kirby
University of Newcastle

Raven's Progressive Matrices (RPM; Raven, 1938) has become a widely used test of general mental ability. It has been used by Cattell (1971) as a measure of fluid intelligence and by Jensen (1970) as an index of Level II ability. It is also commonly used in education as a test of general intelligence or nonverbal reasoning. Historically the test seems to have emerged as an attempt to assess Spearman's *g*, "the eduction of relations and correlates", - and while the concept of general intelligence still has broad support (e.g. Cooley, 1976), there are increasing efforts to understand it in terms of more discrete and more easily defined cognitive processes (e.g. Carroll, 1976; Das, Kirby & Jarman, 1979; Hunt, 1976). Recently, part of this effort has been specifically directed at describing the processes which underlie successful solution of RPM items (Hunt, 1975; Kirby & Das, 1978; Lunneborg, 1977).

Hunt (1975) suggested that for a large number of RPM items, there were two quite different solution algorithms. One was described as the Gestalt algorithm in which one "deals with a problem by using the operations of visual perception, such as the continuation of lines through blank areas and the superimposition of visual images upon each other" (p.133). The Analytic algorithm, on the other hand, "applies logical operations to features contained within elements of the problem matrix" (p.133). Whereas the Gestalt algorithm relies upon the mental manipulation of sensory images, the Analytic algorithm deals with abstracted features of the displays, by means of operations such as constancy, supplement/delete, expansion/contraction, addition/subtraction, movement, and composition/decomposition (Hunt, 1975,p.146-147). While the Gestalt algorithm is seen as less developed than the Analytic one (the former being perhaps characteristic of concrete operations, the latter of formal operations), Hunt's analysis demonstrates that use of the Gestalt algorithm alone in Set I of the Advanced Progressive Matrices would result in a score "slightly below average performance in the normal adult population" (p.141). Similarly, inspection of the Colored Progressive Matrices (suitable for elementary school children) suggests that the Gestalt algorithm could solve almost all of these items. In all cases, the Analytic algorithm is required for solution of the most difficult items.

If supported empirically, Hunt's analysis could have far-reaching implications. Once the possibility of subjects using different strategies, or algorithms, to solve a particular set of test items is accepted, then interpretations of test performance must become more complex. The total score will not tell us much about the way the individual solved the RPM test items. This leads, as Hunt suggests, to a consideration of an individual's cognitive style in any interpretation of RPM test performance. More specifically it is possible that children who score low on RPM are not spontaneously employing the optimal (i.e. Analytical) algorithm. If such were the case then any assessment of a child's general mental ability would need to consider whether the failure to use the optimal strategy was due to an inability to use it - a mediation deficiency in Flavell's (1970) terms - or because they were simply not aware of the strategy but used it effectively following instruction. The latter position represents a production deficiency as described by Flavell (1970).

What evidence is available for Hunt's algorithm approach? Little direct evidence apart from Hunt's original paper has emerged although there are several sources which provide indirect support. The first of these is a series of item analyses of RPM tests. Corman and Budoff (1974a, 1974b) factor analysed responses on the Colored Progressive Matrices items for a variety of populations. They found consistent support for the existence of four subscales: Discrete Pattern Completion, Simple Continuous Pattern Completion, Continuity and Reconstruction of Simple and Complex Structures, and Reasoning by Analogy. The first three of these would appear to involve components of the Gestalt algorithm, while the last would generally require the Analytic algorithm. However, Corman and Budoff's approach does not map directly onto that of Hunt for he has argued that both the Gestalt and Analytic algorithms could, theoretically, be used on many of the items in Corman and Budoff's first three subscales. For Hunt, item type does not totally dictate strategy use.

Kirby and Das (1978) pursued the relationship between strategy and item type, attempting to demonstrate that those who successfully completed the Reasoning by Analogy items were also more able in the area of analytical (reasoning) ability. Obversely, those who were more successful on the Continuity and Reconstruction of Simple and Complex Structures items were hypothesized to do better on measures of spatial ability. Kirby and Das argued that analytical reasoning ability required similar rule-based Analytical algorithms to the Reasoning by Analogy items, and that spatial ability and the Continuity and Reconstruction of Simple and Complex Structures required similar Gestaltic algorithms. If these hypotheses had been correct, Kirby and Das would have found that their measures of spatial ability would have correlated more highly with the Continuity and Reconstruction subscale than with the Reasoning by Analogy subscale. Similarly, analytical reasoning ability would have been correlated more highly with the Reasoning by Analogy subscale than with the other. In a sample of nine year old children, these hypotheses were not found to be tenable. While this result does not disprove Hunt's position, it does not offer it the support it might have.

Lawson and Kirby (1978) examined the role of early items in the development of the two solution strategies. In the conditions of interest here, their subjects (grade 4, 6 and 8 children) were divided into two groups. Both groups received the standard RPM instructions, followed by six easy items. Then one group (B) did six items which stressed a "good figure" result (similar to Hunt's Gestalt algorithm), while the other (E) did six items that required or at least encouraged analogical reasoning (Analytical algorithm). No special instructions were given and all subjects then proceeded to complete sets C, D and E of the Standard RPM. Though no explicit training was given, this procedure could be referred to as "weak training", insofar as completion of consecutive items of a similar type encouraged subjects to develop a strategy. The two groups of grade 4 subjects, as well as the two groups of grade 6 subjects, did not differ significantly on the Standard RPM posttest; however, grade 8 students who had experienced the weak Analytical training (group E) did significantly out-perform the B group. These results suggest that the grade 8 students had a production deficiency and benefitted from even the slight extra experience of six analytical items. Low numbers of subjects in the grade 4 sample, and the complex nature of sets C, D and E (it not being clear which algorithms were required), leave this study open to other interpretations.

The studies described above were not designed to establish Hunt's alternative solution algorithms, and can only be interpreted as weak support, or nonsupport, for them. Clearly what is needed is a study designed specifically to isolate these two algorithms or strategies.

Identification of strategy use is a major problem. Kirby and Das (1978) used one approach by examining the correlates of performance on RPM problems, arguing that other measures were clearer indicators of the strategies in question. If these strategies are differentially involved, then performance on these tasks should be differentially correlated with performance on the "marker" measures.

A most tempting alternative approach to the isolation of strategies, and a most obvious one, is to question subjects about their reasons for each response and to build-up a classification of strategies from these reasons. While possessing all the difficulties of introspection such a technique is not completely without value. It has been applied recently in investigations of metacognition (see Lawson, in press, for a review) and over many years in Piagetian research with children. The usefulness of the technique clearly depends on the ability of the individual to construct an accurate representation of his cognitive decision-making process. For this reason it seems that the verbal response technique must be used in conjunction with other indices of strategy use.

In using the RPM the researcher is able to examine the pattern of subjects' responses in two different ways, and can relate these to verbal justifications. If items can be identified that clearly require one strategy and not the other, then success on that item can be taken to indicate that the strategy was used. In addition, if particular answer options within an item can be argued to be more Analytic than Gestaltic, then choice of item option can serve as an index of strategy use.

With regard to the RPM, this technique could pick out use of the Analytic algorithm, but not the Gestaltic, as all items soluble by the Gestaltic algorithm are also soluble by the Analytic. As item difficulty is also greater in the problems presumed to require the Analytic algorithm, this implies that higher total score is related to use of Analytic strategy.

The final technique to be discussed employes strategy training. If the different algorithms exist as different strategies, instruction in use of a particular one could increase use of that algorithm. This increased use of a particular algorithm could be detected by analysis of both justifications and response patterns. Because the algorithms seem to be related to ability, and even to developmental level, it is by no means certain that subjects can be taught to use particularly the Analytic algorithm more effectively.

The purpose of this study is to attempt to identify Hunt's two processing algorithms as solution strategies in RPM items, by means of the techniques described above. More specifically our aims are to (1) see whether two distinct strategies can be isolated using both verbal justifications and response patterns as indicated, (2) see whether these strategies can be trained, (3) investigate how strategy use is related to performance both overall and on specific items, and (4) to relate these results to Hunt's analysis. Because the algorithms may be related to developmental stages, subjects of an age at the beginning of formal operations were most appropriate for the study.

Method

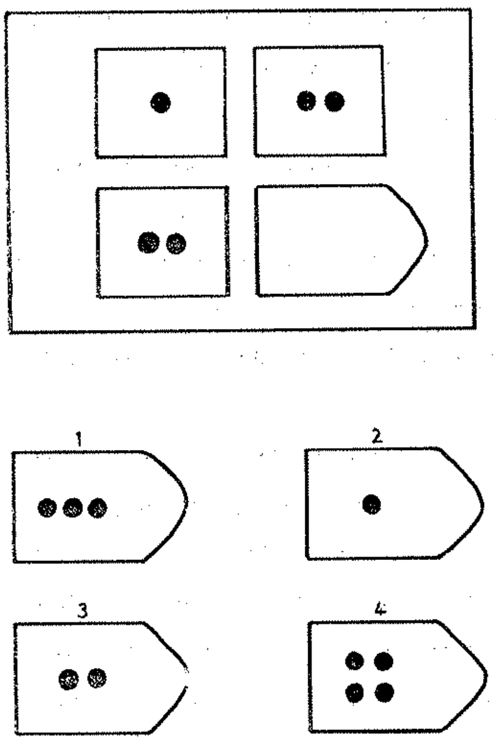
Subjects

The subjects were 30 grade 6 boys attending a metropolitan primary school in Adelaide. The boys came from each of the school's grade 6 classes which are not streamed on any systematic basis. The mean age of the boys was 125 months with a standard deviation of 5.8 months. Subjects were randomly assigned to either of the two training groups.

Training. One set of instructions (the G instructions) was designed to train the use of the Gestalt algorithm, while the other (A) instructions were designed to train use of the Analytic algorithm. In the G instructions subjects were told: "We are going to do problems in which you have to work out what is missing from a picture. Each problem is basically a pattern with a piece missing. You have to pick out a piece to put in that space so that the picture or the pattern is finished. Look at each of the pictures and try to pick out the missing pieces. You have to pick the piece that completes the picture - that makes it look like a good pattern." Following this general introduction subjects were taken through items A1, Ab7, A7, and Ab12 from the Colored Progressive Matrices. On each of these items the above training was emphasized and subjects were given feedback after each item to ensure that they had used the appropriate strategy. Subjects were then given two specially constructed ambiguous items (see Figure 1 for an example). Again use of the G strategy was stressed and choice of the G options emphasized (option 3 in Figure 1 completes the pattern, balances the picture).

The same format was used for the A training with appropriate modifications. The general instructions to subjects indicated that: "We are going to do problems in which you have to work out what is missing from a picture. For each problem there is a rule which tells you what should be in the empty space. What you have to do is work out the rule and then work out what the missing piece is. First try to work out the rule and it will help you work out what the missing piece is." For the A group the practice items used were A1, B3, B9, and Ab12 from the Colored Progressive Matrices.

Figure 1. First ambiguous training item



For the two ambiguous items (the same items used with the G group) choice of the analytic options (e.g. 1 or 4 in Figure 1) and of the analytic strategy were stressed. Both forms of training lasted approximately 10 minutes.

Posttest measures. Following the completion of the training session subjects were given the posttest items. These sets of items were the same for both groups. First were four more ambiguous items which were designed to provide a check on strategy use following training. Then came items 1 - 10 from Set 1 of the Advanced Programme Matrices. These items had been analysed by Hunt (1974, Table 6.1) who suggested that while items 1 - 6 could be solved by use of either algorithm, items 7 - 10 required use of the Analytic algorithm for solution.

Following completion of each of these items subjects were asked to say why they had chosen a particular option.

Procedure

All training and testing was carried out individually by one experimenter. Subjects indicated their choice of options on an answer sheet and their verbal justifications were taperecorded during performance of the Set 1 items. The complete training and testing session lasted approximately 40 mins for each subject.

Coding of Justification Responses

Following the completion of data gathering the subjects justification responses were transcribed for analysis. Each response was then classified using a three part coding system. Justifications were classified as G if subjects stressed any of the following factors: the spatial nature of the array; the continuation of elements of the array; "coloring-in" of parts; or the visual appearance of the chosen option (e.g. the balance or symmetry or "It's getting bigger as it goes across", "It goes skinny then a bit wider"). Justifications which stressed rules, such as: a counting rule ("this one has 3, this one 2, so this must have 1"); a position rule ("top left, top right, bottom left, so it must be bottom right"); or a subtraction rule, were classified as A.

All justification responses were classified by two raters, neither of whom knew the subjects' training group membership. After discussion of sample responses both raters agreed on virtually all responses. If no agreement could be reached about a response, or the response was tautological or irrelevant ("Because its got all little squares in it and there's no other like it"), it was placed in the Unclassifiable (U) category.

Results

Ambiguous items. Frequencies of A or G responses for the four ambiguous items are shown in Table 1. It can be seen that those undergoing G training responded almost exclusively in a G fashion, i.e. they chose options which completed or balanced the pattern in the problem matrix.

TABLE 1: Frequency of Gestaltic or Analytic responses in four ambiguous items

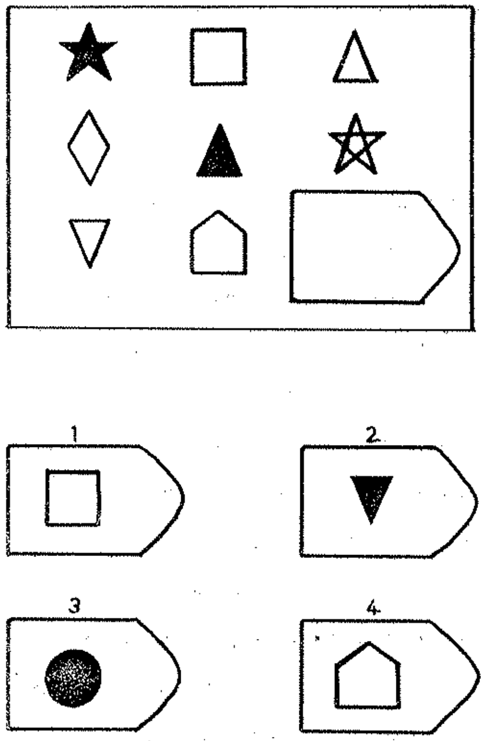
	Item									
	1		2		3		4		overall	
Training Group	G*	A	G	A	G	A	G	A	G	A
Gestalt	15	0	15	0	15	0	11	3	56	3
Analytic	0	15	4	11	5	9	13	1	22	36

* G indicates number of choices of Gestalt options, A indicates number of Analytic options. N = 15 in each training group. Some options were appropriate to neither algorithms; thus all cells do not add to 15.

The pattern of responses in the A training group is less straightforward. Overall there is still a tendency for predominantly A responding in the A group, certainly much more so than in the G group. The difference between training groups in the overall patterns of response is significant. (x^2=40.2 p<.001).

However, choice of A options in the A group weakens in the second and third items, and almost disappears in the fourth item. It would appear that while both forms of training have effectively established different strategies, the A strategy is quite fragile, disappearing when difficult problems are encountered. Item 4, for which G responding was dominant in both groups is shown in Figure 2.

Figure 2. Fourth ambiguous test item



For 10 year old children this is clearly a very difficult problem to process analytically, for the very powerful diagonal of black figures must be ignored. In this item option 2 is the analytic response (having four points), and either 2 or 3 can be regarded as Gestaltic (being colored in). Given the difficulty of the analytic rule it is perhaps not surprising that the A group subjects "regress" to G responding for this time.

Set 1 items. Table 2 contains the frequencies of correct and incorrect responses for each item for the two training groups. Each justification response is classified according to the coding system described previously.

There are very few differences in the number of correct responses between the groups. Only on item 10 is there any major disparity in the level of performance. For the first nine Set 1 items as a whole neither type of training appears to offer subjects an advantage. In part this would be expected from Hunt's (1975) analysis, for he argues that the first six of these items can be solved using either algorithm.

Table 2 : Frequency of correct and incorrect responses on Set 1 items, classified by types of justification and training group.

GROUP

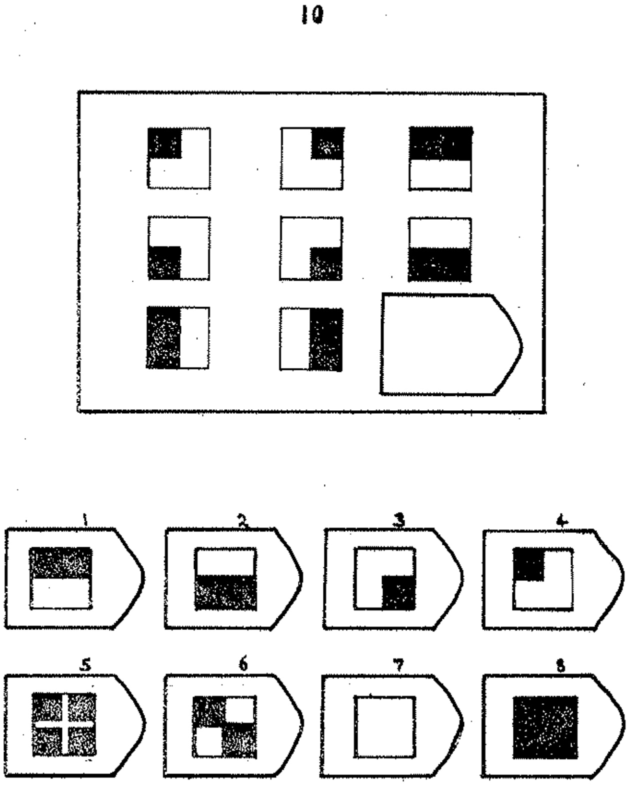
Set 1 Item	Gestalt Training				Analytic Training		
	Justification*	Correct	Incorrect		Justification	Correct	Incorrect
1	G	13	1		G	12	
	A	1			A	3	
2	G	6	1		G	2	
	A	8			A	13	
3	G	12			G	12	1
	A				A		1
	U	1	2		U	1	
4	G	14			G	15	
	A				A		
	U	1			U		
5	G	6	1		G	8	1
	A	1			A	1	
	U	5	1		U	2	3
6	G		1		G	1	2
	A	10	3		A	8	3
	U	1			U		1
7	G		1		G	3	1
	A	10	2		A	6	
	U	1	1		U	1	4
8	G		4		G	2	2
	A	5	1		A	5	2
	U		5		U		4
9	G	1	2		G		2
	A	3	9		A	4	8
	U				U		1
10	G	12	1		G	3	2
	A				A	3	5
	U	2			U		2

*G indicates a response classified as Gestaltic, A as analytic, and U as unclassifiable.

However, in these results the degree of success in solving items 7, 8, and 9 is much the same for both groups, and this is not what would be expected on the basis of Hunt's analysis. If in fact these three items do require the use of the analytic algorithms then many of the subjects in the G group do seem able to switch to that algorithm despite G training. The justification data seems to reinforce the this interpretation for from Table 2 it is clear that the frequency of A justifications given by the G group does increase markedly on items 6 - 9.

Item 10 produced a most interesting set of responses. In spite of Hunt's classification of the item as Analytic, it can be seen that the Gestalt-trained group was the more successful. More can be learned when subjects justifications of their choices on this item are examined. In the G group the justifications are almost exclusively Gestaltic; in the A group they are more mixed, though A reasons predominate. However, for the A group a greater degree of success attended G justifications ($^{3}/5$) than did A justification ($^{3}/8$). Inspection of item 10 (see Figure 3) and of the typical G justification reveals why this item is approached in this manner. Subjects who were correct apparently ignored most of the display and concentrated their attention upon the last row or column (e.g. "As it goes down it gradually fills up until it gradually fills right up"). Other G justifications indicated a type of superimposition operation: "Put the first and second ones together you get a half ... so when you put the two halves together you get a whole". Thus, contrary to Hunt's view, it seems that item 10 is soluble using the Gestalt algorithm, specifically by a variation of the superimposition operation.

Figure 3. Item 10 of Set 1.



A final point of interest in Table 2 concerns the relationship between training group membership and justification type. Training group membership does not always accurately predict the type of justification offered by subjects. This is apparent in the case of Item 2 for the C group and items 1, 4 and 5 for the A group. In these cases the nature of the item seems to be a more potent factor than training.

Discussion and Conclusion

This study was designed to address four issues: (1) can two distinct strategies be identified? (2) can the strategies be trained? (3) is strategy use related to performance? and (4) can these results be related to Hunt's anaylsis? We will discuss each of these issues in turn.

Can two strategies be identified? It would appear that the answer to this question is Yes. It was not difficult to follow Hunt's analysis and devise two brief training sessions. Similarly it was not difficult to unambiguously assign subjects' responses to categories. Subjects' choices in three of the four ambiguous items and in item 10 of Set 1 were different according to group membership. The patterns of response on these items are especially significant for they show that strategy is not wholly dictated by item type. Subjects given different sets of training instructions do respond in identifiably different ways.

The validity of our arguments for strategy identification depend of course upon the criteria we have used. The results here would give reasonable support to the usefulness of using two main indices in identifying a strategy. The first, the ambiguous items, did provide a way of checking on the influence of training instructions and at least for the first 3 of these items training group is a good predictor of response. To this extent training does seem to have encouraged subjects to adopt different strategies. The use of the ambiguous items allows analysis of the interaction between two components discussed in the introduction to this paper - training, and analysis of the subject's pattern of response. Because of this the ambiguous items provide an advantage over the normal RPM items in which the design of the options for an item is often less than optimal for diagnosis of strategy use. It is often impossible to make anything of errors on RPM items which would provide clues as to why a subject was not successful.

The justification responses were used as a second source of information about strategies, and as indicated in the analysis of the data presented in Table 2, they do appear to elucidate strategy use on particular items. When subjects' justifications of responses are taken into consideration the interaction between training and item-type can be examined more closely. Thus it appears from Table 2 that there is a considerable degree of switching between A and G strategies by subjects within training groups. The justifications given by the G group for their choices on items 6 and 7 (see Table 2) are good examples of this strategy switching. On these items the majority of reasons given by G subjects were of an analytic nature. However the strongest support for the use of justification data emerges from an analysis of responses in item 10. For this item where the present results are in conflict with Hunt's analysis, the justifications given by subjects provide a possible means of resolving the conflict.

It appears that subjects strategies did not persist in difficult problems (the fourth ambiguous item) or in problems which seemed to have a bias toward a particular strategy (Item 1 in Set 1). However, this is not surprising and in no way weakens the case for identifying the two strategy types. Subjects do not begin such tests with a total lack of knowledge about how to do them, and in any case would, from experience with the earlier items, learn ways of dealing with them. This is clear from the "weak training" results obtained by Lawson and Kirby (1978).

<u>Can the strategies be trained</u>? Again the answer would be a qualified Yes. Subjects did behave differently and in the predicted directions following training. The training sessions were quite brief, but attempted to follow Hunt's descriptions of the two algorithms. Longer training might well produce a stronger effect, though it would be difficult to overcome two biases: the apparent developmental trend from G and A thinking, and the dominance of the item type in problem solving. The only evidence of the developmental trend in the present study is the performance of the A groups on the fourth ambiguous item. Upon encountering a difficult item they appeared to have "regressed" to a G strategy.

Item bias is also quite strong. If items have elements that vary in an obvious numeric fashion (e.g. three dots, then two dots, then one dot) it is difficult to imagine a way to stop ten year old children from counting and arriving at simple rule. This is particularly so because the Analytic algorithm provides answers which are in agreement with those of the Gestalt algorithm. Thus a child trained Gestaltically may still be attending to numerical rules during training; when able to respond freely, his mentioning of a numerical rule would have his response classified as Analytic.

The influence of item type is apparent when subjects' justifications of their responses are considered. If justification does index strategy then it appears that subjects are quite flexible in using different strategies even though they have received special training at the start of a testing session. Since subjects might well be switching between alternative strategies, the influence of training instructions must not be overestimated. In this regard the subjects' justifications are an important aid in interpreting the effects of strategy training.

Is strategy use related to performance? This is an important issue which has not received adequate attention in the strategy literature. Table 2 above shows clearly that not only are there preferred strategies for each Set 1 problem, but also that a particular strategy (G for items 1, 3, 4, 5 and 10; A for items 2, 6, 7, 8 and 9) is overwhelmingly successful in each case. In item 10 the preferred strategy (A in the A group) is not the more successful one. In this case either training handicapped the A group, or helped the G group considerably. Appropriate strategy use is related to successful performance; only in item 10 did strategy training appear to dominate item based strategy bias.

Relation to Hunt's classification. While these results confirm Hunt's analysis to a great extent, they do suggest some modifications. Item 10 provides the major point of difference: our results suggest that this item is soluble by the Gestalt algorithm and that subjects may be using operations such as a variation of the superimposition one described by Hunt.

Both items 2 and 6 also depart from Hunt's analysis. While they may both be soluble by the Gestalt algorithm, only a minority of subjects in item 2 and only one subject in item 6 (of those who were correct) did use that strategy. Both items have elements that have small numbers of dots or lines, and subjects generally counted these and formed a rule.

There is also weak support in these data for Hunt's contention that the algorithms are developmentally related. In the fourth ambiguous item, subjects trained analytically did switch to a Gestaltic strategy, even though they later returned to an analytic one. The A group also did not have an advantage in overall performance on Set 1, if anything a disadvantage. While this does not support a production deficiency interpretation, it could be understood if the age of twelve were not thought of as a likely point for the beginning of formal operations. Evidence of a production deficiency might be found if older subjects were used.

Conclusions. It would be an understatement to describe the RPM as a complex test. Hunt has shown there are two types of strategies or algorithms involved in the different items. In a previous study (Lawson & Kirby, 1978) we showed that simply manipulating the early items done could change some subjects' performance. The present study has experimentally verified the existence of the two strategies, and shown that they are to a certain extent subject to training. The individual items, however, were found to have a powerful effect upon the strategies subjects adopted. Thus both training and item type influenced the strategy chosen, which in turn was vitally related to subjects' level of performance.

References

Bridgeman, B. & Buttram, J. Race differences on nonverbal analogy test performance as a function of verbal strategy training. Journal of Educational Psychology, 1975, 67, 586-590.

Carroll, J.B. Psychometric tests as cognitive tasks: A new structure of intellect. In L. Resnick (Ed.) The nature of intelligence. Hillsdale, N.J.: Erlbaum, 1976.

Cattell, R.B. Abilities: Their structure, growth, and action. Boston: Houghton Mifflin Co., 1971.

Cooley, W.W. Who needs general intelligence? In L.B. Resnick (Ed.) The nature of intelligence. Hillsdale, N.J.: Erlbaum, 1976.

Corman, L. & Budoff, M. Factor structures of retarded and nonretarded children on Raven's Progressive Matrices. Educational and Psychological Measurement, 1974, 34, 407-412 (a).

Corman, L. & Budoff, M. Factor structures of Spanish-speaking and non-Spanish-speaking children on Raven's Progressive Matrices. Educational and Psychological Measurement, 1974, 34, 977-981 (b).

Das, J.P., Kirby, J.R. & Jarman, R.F. Simultaneous and successive cognitive processes. New York: Academic Press, 1979.

Flavell, J.H. Developmental studies of mediated memory. In L. Lipsitt & H.W. Reese (Eds.) Advances in child development and behavior. Vol. 5, N.Y.: Academic, 1970.

Hunt, E. Quote the Raven? Nevermore! In L. Gregg (Ed.), Knowledge and cognition. Potomac, Md.: Lawrence Erlbaum, 1975.

Hunt, E. Varieties of cognitive power. In L. Resnick (Ed.), The nature of intelligence. Hillsdale, N.J.: Erlbaum, 1976.

Jensen, A.R. Hierarchical theories of mental ability. In W.B. Dockerell (Ed.), On intelligence. Toronto: Methuen, 1970.

Kirby, J.R. & Das, J.P. Skills underlying Colored Progressive Matrices. Alberta Journal of Educational Research, 1978, 24, 94-99.

Lawson, M.J. Metamemory: Making decisions about strategies. In J.B. Biggs and J.R. Kirby (Eds.), Instructional processes and individual differences in learning, in preparation.

Lawson, M. & Kirby, J.R. An initial investigation of the development of solution strategies for Raven's Progressive Matrices items. Unpublished manuscript, Flinders University, Australia, 1978.

Lunneborg, C.E. Information processing correlates of Raven's Progressive Matrices tests. Report 77-10, Educational Assessment Center, University of Washington, 1977.

Raven, J.C. Guide to the standard progressive matrices. London: H.K. Lewis, 1938.