

79 REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 18 AFOSR TR- 78-0131 ✓	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) 6 THE SPY RING HISTORY TEST. 2	5. TYPE OF REPORT & PERIOD COVERED Interim Scientific 1 Aug 75 - 31 Jul 76	
7. AUTHOR(s) 19 Gordon/Pask	6. PERFORMING ORG. REPORT NUMBER Research Note, No. 1.	
9. PERFORMING ORGANIZATION NAME AND ADDRESS System Research Ltd., Woodville House 37, Sheen Road, Richmond, Surrey, England 17A2	8. CONTRACT OR GRANT NUMBER(s) 15 F44620-76-C-0003	
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research (NL) 16 Bolling AFB, DC 20332	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 16 2313A2	
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	12. REPORT DATE 31 Jul 76	
	13. NUMBER OF PAGES 11 1274 p.	
	15. SECURITY CLASS. (of this report) Unclassified	
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE	
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) The "Spy Ring History Test" is described. It is one of several, similarly structured, mental tests intended to discriminate between individual cognitive styles, and individual tendencies to learn according to different plans and strategies for exploring (or assimilating) subject matter. Form II (copies of which accompanied Progress Report No. 1) is typical and may be referenced as representative. The tests are not meant to determine absolute levels of performance, learning rate, or retention, though there may be (and, by → (cont. over))		

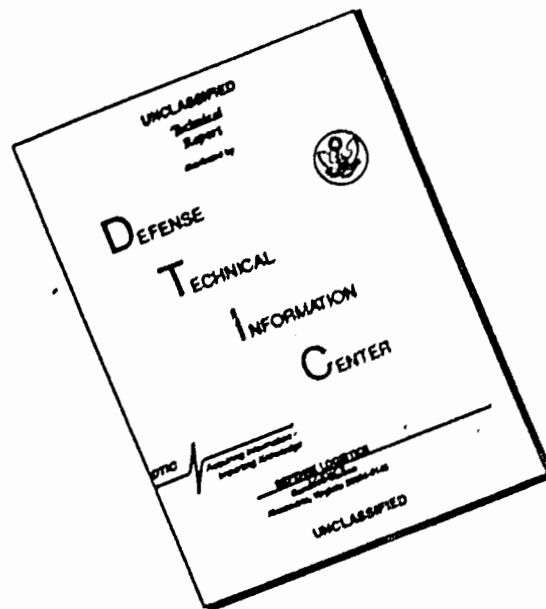
AD A 050458

DDC FILE COPY

DDC
PROCESSED
FEB 23 1976
F

339930

DISCLAIMER NOTICE



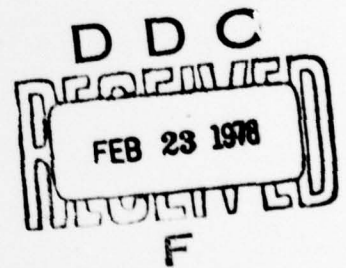
THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

RESEARCH NOTE, NO. 1.

"THE SPY RING HISTORY TEST"

Contract No F44620-76-C-0003

Date: 1 August 1975 - 31 July 1976



System Research Ltd
2 Richmond Hill
Richmond
Surrey
England

01 940 0801/5025

Approved for public release:
distribution unlimited.

20. ^{Cont} hypothesis, there are) relations between test scores indicating stylistic differences, and a respondent's performance in specific tasks.

The dimensions which have emerged as significant discriminators of style and strategy (from studies of educationally sized learning) refer to quite specific cognitive mechanisms and are most readily observed as features of continuing behaviour and cognition. They are believed to have special relevance to such inherently complex and individualised activities as hypothesis formation and system design.

A

ACCESSION for	
NTIS	Write Section <input checked="" type="checkbox"/>
DDC	Buff Section <input type="checkbox"/>
UNANNOUNCED	<input type="checkbox"/>
JUSTIFICATION	
BY	
DISTRIBUTION AVAILABILITY CODES	
Dist	SPECIAL
A	

1473B

Research Note. No 1: The Spy Ring History Test

Index

	page
1. Background	1
2. Mental Operations relevant to complex learning	1
3. Method of Measurement	2
4. Discussion and Criticism of the Method: Validity and Reliability	4
5. Formal Structure of topics in the tests	5
6. An Interpretation for "The Spy Ring History Test"	6
7. Differential Response Scoring	7
8. Isomorphic Tests, Matched Tests, and More or Less complex tests	8
References	9

1st August 1975-31 July 1976

The "Spy Ring History Test" is one of several, similarly structured, mental tests intended to discriminate between individual cognitive styles, and individual tendencies to learn according to different plans and strategies for exploring (or assimilating) subject matter. Form II (copies of which accompanied Progress Report No. 1) is typical and may be referenced as representative. The tests are not meant to determine absolute levels of performance, learning rate, or retention, though there may be (and, by hypothesis, there are) relations between test scores indicating stylistic differences, and a respondent's performance in specific tasks.

1. Background.

Individual differences in learning strategies and information processing have been revealed in numerous studies, several of which are referenced in a paper (Pask 1976b) that also summarises the results obtained by this research group in the laboratory and in the field. Some differences (for example, on Kagan's "impulse/reflective" dimension; on Witkin's "Field Dependent/Field Independent" dimension and differential scores on aptitude tests, such as the battery underlying Guilford's "Structure of Intellect" model) are consistently related to age, sex, personality, etc., and may change in a predictable manner during development. These indices are, however, relatively gross because of the underlying distinction, or because the differences are quantified (in terms of response to small scale and substantially independent questions, problems, situations, etc.), or both. Moreover, these stylistic distinctions are not primarily intended to uncover detailed cognitive mechanisms, or to bear directly upon lengthy stretches of behaviour.*

2. Mental Operations relevant to complex learning.

The dimensions which have emerged as significant discriminators of style and strategy (from studies of educationally sized learning) refer to quite specific cognitive mechanisms and are most readily observed as features of continuing behaviour and cognition. They are believed to have special relevance to such inherently complex and individualised activities as hypothesis formation and system design.

During this type of learning it is possible to distinguish mental operations that are constituents of either understanding (used, in a technical sense, (Pask 1976a) to mean the acquisition of a stable and reconstructible concept) or deep level processing (used in the same symposium by Marton and Saljo 1976 as distinct from "surface level" processing). There is good evidence that the retention of useable (in contrast to parroted) concepts, and creative performance in problem solving, etc., based on this knowledge, depends upon an understanding of background topics or (perhaps equivalently) that information about these topics is deep level processed.

* Clearly, there is an indirect relation, for example, Field Dependent individuals are likely to make better salesmen and Field Independent individuals better formal thinkers. However, these predictions are molar; for instance, the Field Dependent Individual is more likely to notice cues in his environment, including personal gesture, voice inflexion, etc., but without commitment to how these cues are used as inputs to salesmen's performance.

In turn, the achievement of understanding depends upon an interlocked series of description building (goal specifying) operations, and procedure building (algorithmic or rule constructing) operations, while an individual is coming to grips with background data and principles. Under standard conditions (for details see Pask 1976a) it is fairly easy to ensure that any individual combines appropriately both kinds of operation and reaches an understanding of salient topics: if necessary, by providing external information and assistance to augment his mental repertoire. It is also practicable to "titrate" the amount and type of assistance (if any) that the individual needs in order to reach understanding.

Without control, individuals show definite and, it appears, reliably transferable biases. Some habitually build descriptions, which may (or may not) hang together, but are often unable to build procedures which realise what is described. In the "Spy Ring History Test", these individuals are classified as "Comprehension Learners". Others, habitually build procedures, provided a description is given, or a goal is stated, (if not, they tend to learn a collection of unrelated rules). These individuals are classified as "operation learners". Among "Operation Learners" are those who have a bias towards learning rules that are given (for example, paired associate lists) and others who will discover a rule that generates the list.

Finally, some individuals understand topics without prompting; they build both descriptions and procedures and integrate them. Such respondents have appreciable comprehension learning scores and operation learning scores, and it appears that at least some of their operation learning is due to rule discovery. They are classified as "versatile".

On theoretical grounds, description building is expected to vary in kind from the construction of descriptions that relate many variables (or properties) to precise descriptions that involve only a few. The expected tendency has been observed, and appears to be individually specific; in the "Spy Ring History Test", the scoring scheme is designed to pick up a "Global/Local" tendency. A further prediction is that respondents will be more-or-less inclined to invent formally redundant descriptive properties which they use in memorising or problem solving. Such tendencies are indicated by many indices of divergent thinking and creativity, but have also been observed as differential components of performance; for example, in taxonomy recall (Pask and Scott 1972) where the proposed discrimination of "redundant/irredundant holists" is based upon the prominence of such behaviours. An index of inventiveness (meaning "the invention of descriptive properties for use in recall") is embodied in the "Spy Ring History Test" scoring scheme. As an hypothesis, "inventiveness", in this sense, should correlate positively with the "rule discovery" type of "operation learning"; with readiness to structure an unordered environment in a personal but coherent manner (one component of "learning to learn" noted in Pask 1976b) and also with Witkin's Field Independence. These possibilities are being investigated.

3. Method of Measurement

Any method able to detect stylistic differences, of the type noted in the previous section, must, by definition, sample stretches of behaviour large enough to provide evidence for the underlying processes found to exist (by independent studies) and to have an important bearing upon educationally sized learning and problem solving.

The preferred method (noted in the last section) is to hold the respondent in controlled conditions whilst he learns a large and well structured body of knowledge and to insist upon a standard requirement of understanding concepts by eliciting explanations in response to "how" and "why" questions. Either disciplined interviews, or mechanised systems, may be used for this purpose (the cognitive operations under discussion were initially uncovered and formulated in this context). By their nature such experiments are time consuming and the explanatory responses, even if elicited in a mechanical (and objectively interpretable) format, are expensive to score.

Several compromise methods have been considered and experimentally tried.

(a) Existing test items with known inter-item consistency (Witkin's (1962) "Embedded Figures" test, and "Closure Figures", Thurstone, (1950) are administered, as usual, to form a series of a-priori independent trials. The respondent is asked how and why he responded as he did, in an attempt to reveal mental mechanisms (instead of eliciting a dichotomous response in a fixed interval after each stimulus, or using a power test technique). There is an obvious difficulty over interpreting his response. However, for individuals willing to state reasons, it seems as though people adopt different but self consistent styles, and that mental operations similar to those of the previous section are at work in their heads. (For example, the responses suggest a complex type of Field Dependence/Independence - like the traits discussed in Witkin's latest report (1975) - and offer an explanation for the variability of standard-form test-scores). Relatively rapid administration is possible, but data evaluation is very expensive and the validity of the test is questionable. True, solving stimulus problems is a measure of Field Dependency but under these circumstances Field Dependency appears as a complex quantity and the lack of logical connections between test items makes it difficult to identify the test situation with a real life task.

(b) The respondent learns a sizeable body of knowledge presented as a series of arbitrarily, and usually linearly ordered, data sets. For each data set, he reaches a criterion-correct response level, indicating that data is stored (for certain) in a particular manner (usually as paired-associate lists). In fact, the concepts to be learned have a coherent and known-to-the-designer-structure (they are organised in a mathematically quantifiable manner as topics in an entailment structure (Pask, 1976a, for example). Further, all of the organising principles could be discerned by the respondent, since the instructions and task description are sufficient to permit unambiguous inference of the structure that exists. Given appropriate inferences, the stored (retention criterion checked) data can be reorganised in memory.

The task is chosen so that interference between the prescribed data sets will grossly overload a respondent who does not transform the memorised information (though several types and degrees of transformation are possible and, depending upon the individual, might be equally effective). Subsequently, recall of and predictive extrapolation from the learned material is elicited by fixed questions (mostly requiring constructive but

objectively scorable responses; not, as in the "preferred" method by explanations). These questions are interdependent and are directly related to the known structure of the subject matter so that it is possible to detect the use of specific cognitive operations.

Unfortunately, administration is time consuming; there is some difficulty in securing the initial learning criteria and, although discrimination is possible, less information is obtained than under standard (understanding) conditions. The "Spy Ring History Test" is a test of this type.

4. Discussion and Criticism of the Method; Validity and Reliability

The possible (and realistic) criticism that tests of this kind are uneconomic can only be countered by commenting that, so far as we know, it is impossible to obtain something for nothing! Serious attempts have been made to derive stylistic discrimination indices using objectively scorable responses to small items, questions and problems, but with unsatisfactory results. Moreover, there is a gradation; the less complex the task, the less discriminating (roughly speaking) is the data.

A converse criterion is also legitimate. If a complex task is used then more data might be obtained by replacing the correct response criteria for retention of the original data sets by an index of understanding, and using explanation-demanding "how" questions to elicit overall recall. Pilot studies (both in this and other laboratories) show rich and interesting explanatory responses; so far as possible, the readily scored questions have been adapted to fit the known possibilities.

The computer administered form of the "Spy Ring History Test" (noted in Progress Report No. 2, and to be described in Progress Report No. 3) is a step in the direction of obtaining depth information about style without relinquishing objectivity as well as a device for standardising the retention criteria for the initial data sets.

The validity of a test of this type is hardly in doubt. On the one hand, by dint of the underlying structure of topics and questions, the test responses depend upon specific and exteriorised cognitive operations. If differential responses are at all discriminating, then these are the mental operations discriminated. On the other hand, it is not difficult to place the highly structured, test-task in correspondence with specific aspects of any similarly structured real life task, and to argue, on firm grounds, that if the test responses are at all predictive, then they predict an individual's behaviour in situations that occur repeatedly in the course of learning, creative activity, and large scale problem solving. So far as task transfer validity is concerned, there are obvious advantages in using a complex and well structured test situation.

The reliability of such a test is a very different matter. Both inter-item and split-half reliability determination is excluded because the test problems are not homogeneous and the test questions are necessarily interdependent. Since respondents "see through" (ie. see the "structure of") the topics in the test task, no meaningful test/retest reliability index is obtainable. Studies have been carried out in which different forms of the test, based upon the theme of "A Spy Ring History", but varying in the number of spies, the complexity and stability of the spy ring network, etc., have been administered to the same respondents. Even though a practice session is built into the test administration, a

significant improvement in overall performance is evidenced. But, unless there is intervening training specifically designed to modify style, the differential scores are individually consistent (calculated only for the main indices, comprehension learning, operation learning and versatility) and, in some cases, they become "peaked" or differently "polarised".

Summary data for 52 subjects is as follows : rank correlation coefficients for the three differential indices are

- (a) comprehension learning - operation learning 0.82.
- (b) comprehension learning - versatility 0.70
- (c) operation learning - versatility 0.76.

The values are significant at the .1% point ($p < 0.01$, one-tailed test).

Because the test task is highly structured it is easy to generate matched alternative forms of comparable difficulty*, and several are now available. Although one might have greater confidence in reliabilities, or consistency indices based upon readministration of matched tests of similar difficulty (rather than matched tests of different difficulty as in the last paragraph) the data so far available is too sparse to warrant more than the "intuitive" comment that individual differential test scores "appear to be consistent between sessions".

5. Formal Structure of topics in the tests.

Any stylistic test of the "Spy Ring History Test" family is characterised as follows :

(1) There are \underline{n} rules, G_j , $j = 1, \dots, n$, $n + 1$ determining the permitted behaviours of \underline{m} elements or entities of \underline{r} ($\underline{m} \approx \underline{r}$) types, each type being determined by one (entity specific) rule. The \underline{n} rules are conveniently represented (to the test designer) as non-deterministic automata or condition-networks; the \underline{m} elements or entities by equivalence classes of states, or by distinguished nodes in the network.

(2) The \underline{n} rules and the \underline{r} types of entity may be specified by \underline{n} distinct lists of at most \mathcal{L} pairs, representing behaviours that may occur according to the \underline{n} permissive rules. That is, both entity type and one permissive rule may be inferred by examining one list; all \underline{n} , by examining \underline{n} such lists.

(3) The entity types are redundantly specified by intensional statements like "this type is a 2 counter" or "that type is an originator".

(4) Some feature common to the \underline{n} rules is redundantly stated by a \underline{p} valued descriptor, which qualifies the actual behaviour lists.

(5) Other features common to the \underline{n} rules are redundantly specified by intensional statements saying "what (conceivably) may take place".

* In fact, the bulk of the work required to generate matched form can be done by a computer program and a previously onerous task is thereby rendered practicable.

(6) There is a periodic functional (or mapping) R , that carries one permissive rule, G into another and repeats at period $n + 1$.

(7) The functional or mapping R , is a function, F , of the structure of the lists noted in (2) (though not, in general, completely determined by them): in any case, given n lists and some of the other information, it is possible to infer the $n + 1$ th rule, either as an extrapolation of $(n, n + 1)$ or an interpolation between the G 's.

(8) At least one semantic interpretation of this formal or systemic information is provided, but no discriminating response is deemed correct because the interpretation is recalled. Any correct response depends, logically, upon formal (or systemic) data and a learner could invent any consistent interpretation of his own, and use it as a legitimate source of recall cues. The score for inventiveness depends upon whether or not the respondent accepts and recalls the given interpretation, or invents his own.

6. An Interpretation for "The Spy Ring History Test".

(a) Considering Form II of this test, $\ell = S$, $m = 5$, $n = 5$, $p = 3$, $r = 3$.

(b) Rules, G_j , of 5(1) are communication networks (directed graphs) showing possible message transmissions between agents or spies G_j : $j = 1, \dots, n + 1$ being the configuration of the network at year or epoch j , in the development of the system.

(c) Each G_j contains the same, individually-name-identified agents (Ajax, Byron, Caesar, Dryden, Euclid). There are $m = 5$ agents $i = 1, \dots, m$ of $r = 3$ types as in 5(3). The types are as follows

$r = 1$; $i = 1, 2$, "Accumulates messages from others and may transmit only after more than one message is gathered".

$r = 2$; $i = 3$ "They send message upon receipt from some other agent but must send before receipt of any further message".

$r = 3$; $i = 4, 5$, "Either originates or transmits messages but does not accumulate messages".

(d) The n rules G_j are specified, as in 5(2), by n lists of $l = S$ agent pairs.

(e) The $p = 3$ value descriptor of 5(4) is "countries". Its values on each agent is "the country in which the agent is located" (Olympia, Ruritania, Transylvania.)

(f) Optionally n graphs G_j may be displayed, in which agent type and the country of location is unspecified (hence, they do not provide as much information as the n lists).

(g) The periods mapping, R , of 5(6) determines the total number of directed arcs in the communication network for a given epoch $j = 1, \dots, n + 1$. These values are

j =	1	2	3	4	5	6
arc number =	6	7	5	6	7	5

(h) Hence, in this case, $G_{j+1} = R(G_j)$ and R is represented as a function of j . It may also be specified as a function, F , of isomorphisms between odd or evenly indexed subgraphs, and as inverse mappings between graphs, thus providing the redundancy called for in 5(7) permitting extrapolation from $j = 5 = n$ to $j = 6 = n + 1$.

(i) The semantic interpretation of 5(8) is provided by background data which the respondent reads before the test and studies as a characterisation of the agents, countries, and activities (including socio economic interpretations to indicate an increase and decrease in the possible communication paths). This background data also embodies the formal, but intensional, specifications noted in 5(3); for example, verbal statements of an agent as someone belonging to a communication cycle, the roles of agents corresponding to the formal types, and so on.

(j) The data sets noted in Section 3(b) are the S member, agent-pair, lists of (d) qualified by "country location". Optionally, (mandatory for group presentation, only) the denuded graphs of (f) may be displayed alongside the lists. Each list is learned in sequence to a criterion (Section 3(b)) of 100% correct response over one repetition (order, pair and qualifiers must all be correct, at this trial stage).

(k) The country "boundaries" are never displayed directly but are easily inferred from the lists, a response, called "map", being elicited. Only n lists are presented as stimuli and the $n + 1^{\text{th}}$ map, graph, or list (for the $n + 1^{\text{th}}$ epoch) is requested, as a predictive response. All other test responses may either be ascribed to inference, or direct recall.

7. Differential Response Scoring.

Deferring discussion of the predictive response, it is possible to construct a complete recall of lists, graphs, maps of all the years (epochs) in many ways. The response formats are chosen as diagnostic of the cognitive style, as reflected in the individual respondent's reconstruction.

After answering questions about country boundaries (draw the map of the countries; place the agents on the map) and questions to determine whether or not the respondent has grasped the intensionally stated rules

(of "being an agent" of "legitimate communication networks" etc.) the respondent is required to reconstruct, as well as he can, all of the possible and actual agent behaviours.

The operation learning score (Section 2) depends upon the inferential question responses and an ability to generate response lists that satisfy the constraints and relations of Section 5 and 6. A bias to learning given rules (Section 2) is evidenced by list recall in strict order, a bias to discovering rules, (Section 2) by the generation of response lists for each epoch that satisfy the constraints - but do not necessarily imitate the particular (S pair) stimulus lists which were presented as representative.

The comprehension learning score (Section 2) depends chiefly upon the "map drawing" response, and constructing map and graph responses for each epoch that preserve the permitted communication paths. At one extreme comprehension learning may degenerate into loose analogising and pattern matching; at the other extreme, using strict analogy relations, it is a powerful method. Since correctness depends only upon formal (as contrasted with semantic or interpretative), recall, the scoring is objective and the degree of accuracy is calculable. The present scoring scheme relies upon simple point assignment (for response components that satisfy formal relations) but an information analysis is possible and promises to yield more discriminating indices.

The specific versatility score is derived from the predictive responses (lists and map) for year (epoch) $n + 1$. Once again the degree of correctness depends upon the number of constraints actually satisfied by the prediction. In the present scoring scheme versatility score points are also given for assimilation of the countries descriptor ("maps", rather than "graph" recall). Generally, the respondent is required to integrate various levels of rule (the G, R and F) in order to obtain a high versatility. As noted earlier, high versatility appears to imply that comprehension and operation learning scores exceed a threshold value, but not vice versa.

Inventiveness (Section 2) is determined by comparing the number of (consistent) invented properties with the number of score points obtained upon questions that ask, directly, about the semantic interpretation given in the background data. The global/local bias (Section 2) was previously calculated by a rough and ready count of responses recalling high/low order relations, but is currently replaced by Decomposition analysis (Atkin 1974), a much more discriminating index.

8. Isomorphic Tests, Matched Tests, and More or less complex tests.

Since the test task is well structured, it is possible to construct isomorphic tests which exactly embody the same formal relations but with a different interpretation. For example, one research group have changed the "Spy Ring History Test Form II" into an isomorph involving schoolboys in place of spies, playgrounds in place of countries, and an interpretation based upon role-dominance as a component in message-transmission. Another independent group are using an isomorph interpreted as a system of chemical molecules.

Since the scoring depends upon formal properties, strict isomorphs are not satisfactory as matched tests. However, matched forms are readily constructed by altering the rules, G, R, F, or requiring interpolative (rather than extrapolative) inference in the predictive (not directly recallable), response. The background (or given) semantic interpretation is also altered to avoid confusion and unwanted interference, or intrusion, effects.

More or less difficult test forms are obtained by systematic variation of the parameters, c , m , n , p , r . The relative information content of these derived forms is calculable and is a rough index of the expected difficulty.

REFERENCES

- ATKIN, R. H. (1974) Mathematical Structure in Human Affairs, London, Heinemann
- MARTON, T. and SALJO, R. (1976a) "On qualitative differences in Learning: I - Outcome and Process", Symposium: Learning Processes and Strategies - I, Br. J. educ. Psychol. 46, 4-11.
- MARTON, T. and SALJO, R. (1976b) "On Qualitative differences in Learning: II - Outcome as a function of the learner's conception of the task", Symposium: Learning Processes and Strategies - II, Br. J. educ. Psychol. 46, 115-127.
- PASK, G. and SCOTT, B. C. E. (1972) "Learning Strategies and Individual Competence", Int. J. Man-Machine Studies, 4, 217-253
- PASK, G. (1976a) "Conversational Techniques in the Study and Practice of Education", Symposium: Learning Processes and Strategies I, Br. J. Educ. Psychol. 46, 12-25
- PASK, G. (1976b) "Styles and Strategies of Learning", Symposium: Learning Processes and Strategies- II, Br. J. Educ. Psychol. 46, 123-148.
- THURSTONE, L. L. (1950) "Some Primary Abilities in Visual Thinking" Report 59, Psychometric Laboratory, University of Chicago.
- WITKIN, H.A., DYK, R. B., FATERSON, H.F., GOODENOUGH, D.R. and KARP, S.A. (1962) Psychological Differentiation. New York, Wiley.
- WITKIN, H. A., MOORE, G. A., GOODENOUGH, D. R. and COX, P. W. (1975) Field Dependent and Field Independent Cognitive Styles and their Educational Implications, Educational Testing Service, Princeton, New Jersey.