SCIENTIA
WNNR · CSIR

SPECIAL REPORT

Rouse

PERS 178

THE DEVELOPMENT OF A HIGH LEVEL SYMBOL GROUPING TEST OF DETERMINATIVE INDUCTION.

PART 1: THEORETICAL CONSIDERATIONS.

NATIONAL INSTITUTE FOR PERSONNEL RESEARCH
COUNCIL FOR SCIENTIFIC AND INDUSTRIAL RESEARCH

CSIR Special Report No. Pers 178 pp. 1-19 U.D.C.159.955.6.072
Johannesburg, South Africa. September, 1972.

The Development of a High Level Symbol Grouping Test of Determinative Induction

Part 1: Theoretical Considerations

by

D.W. Steyn
J.M. Verster

National Institute for Personnel Research
Council for Scientific and Industrial Research

September, 1972.

CSIR PERS 178

0 7988 0250 2 CSIR Special Report PERS 178

Published by
National Institute for Personnel Research,
South African National Council for Scientific and Industrial Research,
P.O. Box 10319,
Johannesburg,
South Africa.

September, 1972.

Printed in the Republic of South Africa by
National Institute for Personnel Research.

Acknowledgements

This project is directed by Mr. D.J.M. Vorster, Director of the National Institute for Personnel Research.

It forms part of the following projects carried out by the Psychometrics Division of the N.I.P.R:

68/18 Construction of a high level Inductive Reasoning test

71/13 Compilation of a high level Differential Test Battery.

The assistance of Mr. M.A. Coulter in conceptualising the project and articulating some of the concepts is gratefully acknowledged.

Introduction

During the past number of years, the N.I.P.R. has evolved a policy of constructing high level selection tests which are based on the conceptualization of human intellect as being differentiated in terms of its constituent abilities.

There is nothing particularly novel about this test construction strategy - it is merely a confirmation of adherence to a particular school of thought as embodied notably in the work of Thurstone[1], (1941), French et al.[2] (1963) and to a lesser extent Guilford[3] (1967).

This approach to test construction was not only motivated by considerations of psychometric elegance and conceptual sophistication, but by the growing realisation that the output and hence the criteria of effectiveness of high level personnel are multidimensional. In high level scientific or administrative work, a large number of abilities come into play and the accurate and unambiguous assessment of these abilities is therefore of considerable importance.

The N.I.P.R. Concept Attainment Test devised by Schepers[4] (1970) is a good example of this strategy. The cognitive domain selected for measurement was clearly defined in theoretical and conceptual terms. The relevant parameters within this broad model were carefully delineated and evaluated against the overall objectives of the test before actual test construction commenced. The end result of this endeavour was a product of such simplicity yet flexibility that the author had no great problems in devising three parallel tests of widely divergent format.

Another more recent example of this type of test construction is afforded by Verster[5] (1972) who started from a precise definition of the type of reasoning he wanted to measure. After having decided on Deductive Reasoning as understood in terms of traditional Aristotlian logic, he then proceeded to make an explicit statement about the parameters of his model by uniquely characterising each item in terms of mood, figure and content. [Verster[6] (1970)]. Again, the one-to-one relationship between the underlying model and the test items is of such an unambiguous nature that the construction of parallel versions of the test becomes little more than a routine clerical task.

It is notable that the authors of these tests make no unsubstantiated claims about their instruments - the models on which they are based provide clear and unequivocal definitions of what is being measured - no more and no less.

The N.I.P.R. has been somewhat less successful in its efforts to devise a test of Inductive Reasoning. During 1969 Barker[7] (1969) devised the Pattern Relations Test (A/15/1). This test was based on the well-known Raven's Progressive Matrices Test under the assumption that it was a measure of induction. This test was appraised by Steyn[8] (1971) who pointed out that Barker's test construction model lacked theoretical substance as a result of her somewhat uncritical acceptance of Raven's rationale. No attempt was made to formulate a formal model for constructing items and the universe of possible culling rules was not defined in any way. A factor study undertaken during 1971 failed to identify the main source of variance of the test and it was recommended that in view of these objections, serious consideration be given to the discontinuation of the test. It was furthermore suggested that the concept induction be re-investigated with the view of constructing an alternative to the A/15/1.

In the following sections the concept of "Induction" will be investigated and discussed against the background of measurable individual differences. The major characteristics of the inductive process will be highlighted and an attempt will be made to synthesize a conceptual model for the development of a suitable measuring device in this broad domain. It is stressed at the outset that the concept 'inductive reasoning' is complex and multifaceted and it is not considered realistic or feasible to construct a test under this blanket term; a somewhat more appropriate approach appears to be the delineation of a specific sub-domain which has relevance to scientific method and which avails itself to the psychometric assessment of individual differences.

Definitions and discussion of induction

English and English[9] (1958) define induction as the "process by which we conclude that what is true of certain individuals is true of a class, what is true of part is true of the whole class, or what is true at certain times will be true in similar circumstances at all times".

This definition of induction which is ascribed to John Stuart Mill is not particularly helpful in developing a model for test construction. It is formulated in such general terms and is so behaviourally sterile that it is only of passing interest to the psychometrician.

Neither does the definition provided by the Shorter Oxford Dictionary (Little, Fowler and Coulson[10]) (1965) throw much light on the topic: " The process of inferring a general law or principle from the observation of particular instances."

It is perhaps more fruitful to contrast induction with deduction. Traditional logic provides a theory of deductive inference which is generally accepted as sufficient "whereas the treatment of induction is often perfunctory and always controversial". (Kneebone,[11] 1963 p. 366). In very general terms however, deduction and induction differ in the sense that in deductive inference the conclusion asserts less than the premises whereas in inductive inference it asserts more. A valid deductive inference is therefore always conclusive, but inductive inferences are inconclusive.

Kneebone[12], (1963) suggests a valuable classification of inductive inferences into two classes namely determinative induction and conceptual induction.

The basis of determinative induction is a statistical principle of optimum choice: from a given range of possibilities, the one that is most appropriate in the light of available evidence has to be selected. It is therefore a selective process of elimination from a prescribed number of alternative possibilities. The reasoning that is involved is essentially of a manipulative kind in the sense that the concepts that are used in the interpretation of the observations are externally given and remain unaffected. Stated in more general terms, it means that the person who makes the inference must first decide what prima facie possibilities present themselves for consideration and then make his choice from among them.

Conceptual induction is of a more fundamental nature. It evolves and defines entirely new concepts to account for observed phenomena and thus transcends the repertoire of previously available possibilities. It is intimately involved in the production of scientific theories such as the General Relativity Theory of Einstein.

In scientific research the two types of induction can be contrasted as follows: Determinative induction comes into play whenever the numerical value of a physical constant or a statistical parameter is estimated on the basis of empirical data or when research effort is directed at the isolation of the specific cause of an observed phenomenon already accounted for by some general theory. It is therefore an almost universal modus operandi of scientific enterprise wherever causal connexions are to be established.

Determinative induction adds factual content to the body of scientific knowledge without necessarily changing its form. Conceptual induction makes its biggest contribution in the area of theory-building where new concepts are attained and formulated which may change the form of the organised body of knowledge and experience.

What is the status of induction in terms of individual differences and psychometric assessment?

French[13] et al.(1963) define induction as "Associated abilities involved in the finding of general concepts that will fit sets of data, the forming and trying out of hypotheses". They then proceed to point out that "factor studies indicate the presence of several factors in this area and some disagreement among investigators, and that it appears impossible to define satisfactorily the several distinct induction factors at this time". (p.19).

This definition is decidedly vague. It does not indicate whether the concepts have to be found from among an explicitly provided set of concepts, in which case determinative induction is measured, or whether concepts have to be attained first before hypotheses can be tested. A careful scrutiny of the tests from French et al.'s[14] Kit of Reference Tests (1963) suggests the latter. In other words, the attainment of certain concepts is left to the testee before he can commence testing hypotheses.

There is a logical dilemma in tests of this nature. A test constructor can never exhaust the universe of inductive rules, and in designing his test therefore limits himself to a certain repertoire of rules.

Let us look at a typical example of an item from a test of this nature:

1. GCPVE 2. KCKOU 3. SCAAK 4. QCQIA 5. UCART

The testee is required to indicate the letter group that does not belong to the other 4. The test constructor had the following rule in mind: The letter group that contains the duplication of the same letter in adjacent positions is the odd-man out. Number 3 fulfills this criterion and is therefore keyed. A particular subject endorses number 5 as being the odd-man out. Let us try and establish why. He carefully scrutinises the letter groups and notices that 5 is the only group starting with a vowel. He has inferred a perfectly legitimate rule and, as a result of the inconclusiveness of inductive reasoning, there is no way of telling whether his rule is any better or worse than the test constructor's rule. Another testee endorses number 1, he has established that this is the only letter group containing only one vowel. A more enterprising testee assigns serial numbers to the letters of the alphabet and then finds the sum of the letter groups. He finds the following totals:

53, 71, 35, 47, 73.

He also endorses number 3, since that is the only total which does not represent a prime number.

The above example should be adequate to illustrate the problems encountered in constructing a test of induction along these lines. It was admittedly an extreme example, but one wonders exactly how many 'ambiguous' items in tests of this nature have gone undetected for long periods of time. The inconclusiveness of this sort of induction could probably be overcome by requesting subjects to write down the rules they followed, but this would involve the test administrator in checking an indeterminate number of rules against the data - hardly a feasible or economical exercise.

From a psychometric point of view, it therefore appears that the dichotomous nature of induction viz. determinative and conceptual should be acknowledged. Measurement of the latter should not be attempted through traditional inductive tests, but could perhaps best be tapped via tests of concept formation and concept attainment. Although Thurstone[15] (1938) and Thurstone and Thurstone[16] (1941) never made a specific statement to the effect, they appear to have been at least partially aware of the dual nature of induction ; the Letter Grouping Test was constructed within a framework of externally given rules which were either explicitly provided or implicitly given in the practice items. A notable feature of the Thurstone induction tests is the liberal use of practice items - in the case of the Letter Grouping Test no less than 14 examples are provided - so that the testee has at least a reasonable repertoire of concepts at his disposal once he starts on the test. It is therefore not unreasonable to speculate that one of the reasons for the successful determination of the induction factor by the Thurstones may be sought in the very way they viewed induction and constructed their tests.

It is clear from the above that the precise nature of induction has not been clearly conceptualized for the purpose of developing an unambiguous test construction model. The following definition is thus offered:

"Given a set of data in which entropy has been reduced by the application of a rule or rules from an explicitly stated set of rules, the act of determinative induction is the quick and accurate identification of which rule most reduces entropy".

Let us examine this definition carefully. It is intentionally stated in information-theory terminology for optimum clarity and precision. The following emerge:

i) The act of determinative induction is clearly stated namely the identification of a specific rule, from a finite set of rules, which accounts best for the order observed in the data. The concept of determinative induction is thus formally differentiated from conceptual induction.

ii) The definition includes the two parameters of the test namely speed of rule identification and accuracy of identification. Individual differences in performance can therefore be expressed in normative terms according to these two conceptually independent dimensions.

The measurement of determinative inductive reasoning ability in the process of selecting high level professional scientists is not only relevant but of great importance. Induction has undoubtedly the most important applications in natural science when general laws are inferred from particular data. Induction always involves a choice, a choice that has to be made as prudently as possible - hence the specification of an accuracy parameter in the test model. Few organizations can afford to employ high level personnel who are capable of impeccable inductive inference at the expense of unrealistic time allocated to the testing of hypotheses. It was therefore decided to specify a speed continuum as a separate dimension of the test.

In the following sections the exact format and nature of the proposed new determinitave inductive reasoning test (Symbol Grouping) will be discussed and its metric characteristics clarified in the light of the definition.

Test Model

The model for test construction was designed to meet the following criteria:

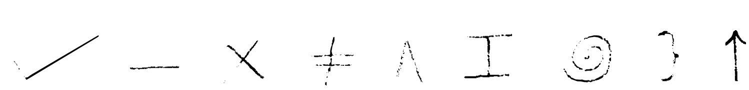
i) The medium in which items are presented should contribute minimally to the variance of the test.

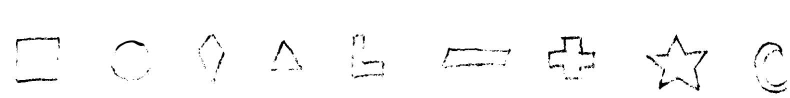
With the aid of smallest space analysis techniques Guttman[17] (1965) was able to demonstrate that a significant proportion of the variance of a battery of 17 of Thurstone's Primary Mental Abilities tests was vehicle variance. This is variance attributable to the various languages of communication employed in the tests. Thus over and above the formal operations which can be built into a test's format to ensure that it is measuring in the desired ability domain, some variance will be due to the fact that the items are presented either pictorially or numerically or verbally as the case may be.

A critical examination of existing tests of induction revealed that both numerically and verbally presented tests contained unintended built-in biases affecting the performance of certain segments of the general population who are in some way deviant in their facility with numbers or letters.

A pictorial presentation using symbols that are universally familiar, in a population which has attained a given level of development, was considered most suited to the requirements of the present test. The symbols, as used in the test will be devoid of operational meaning.

The following set of 18 symbols, most of which were taken from Kellog's[18] (1955) enumeration of universal symbols, were chosen. To provide more scope for generating rules for the test items the symbols were separated into two classes. The two classes of symbols were arbitrarily termed OPEN and CLOSED.

OPEN SYMBOLS 

CLOSED SYMBOLS 

It can easily be appreciated that the perceptual effort required to distinguish between OPEN and CLOSED symbols is minimal.

From a wide range of possibilities only those symbols were chosen which could be considered universally familiar within a western cultural setting and to which everyday verbal tags could be attached readily.

To ensure that no unintended operational signifance is assigned to any of the symbols it will be stressed explicitly in the test instructions that all symbols are to be treated as semantically and operationally devoid of meaning. Each symbol has only two attributes that are of any significance to the test: its unique physical form and its membership to one of the two classes.

ii) The item format should be such that the mental processes required to solve each item are an operational expression of the act of determinative induction as defined.

Certain implications for test construction emerge directly from the definition. The test format should include an explicitly formulated set of rules. The structural format of the items should permit the manifestation of these rules implicitly in an embedding context. Provision should be made for the introduction of rules in either partial or whole form. The format should be such that a subject's response to any particular item provides an unambiguous indication that he has accurately identified the manifestations of the intended rule. Finally, provision should be made for the separate measurement of speed and accuracy dimensions.

To what extent does the sample item cited earlier fulfill these requirements? The item cited is:

1. GCPVE 2. KCKOU 3. SCAAK 4. QCQIA 5. UCART

This basic format is used frequently in various guises in tests of induction. Typically in these tests subjects are told that one of the groups differs from the rest. The task is to indicate the group that is different. A few sample items are usually supplied so that subjects can become familiar with the test before beginning. The tests are usually highly speeded as well.

To what extent does this item type measure up to the format requirements? Since it has already been decided that symbols rather than letters or numbers are to be used the possible problems associated with the use of letters can be ignored. By inspecting this item it can be readily appreciated that its structural format provides a convenient embedding context for the introduction of a wide variety of rules. Provided therefore that the rules to be used are explicitly set out in the instructions, this basic format can be easily adapted to suit our requirements.

The following modifications are proposed. Firstly, the administration procedure is to differ from that which has been traditionally adopted in so called induction tests. The subject faced with a typical test of this type knows that in each item he has merely to detect an 'odd man out' in terms of some criterion. He therefore quickly develops a certain mental set which determines his approach to the test. In the real life situation, however, the research worker or scientist has on the one hand a repertoire of theories about his data (his explicitly given rules) and on the other hand a set of experimentally derived data (analagous to the test item). He has no way of knowing beforehand that one of his available theories will be able to reconcile the findings in his data, neither does he know that his theory will fit all the aspects of his data bar one. The situation is obviously far less structured and the research worker cannot afford to rely on a confident mental set in approaching his data. It is quite possible that more than one of his theories will account perfectly for all the properties in his data. Similarly it is possible that none of his available theories will be adequate to explain his findings.

In an attempt to simulate more closely the real conditions under which determinative induction operates, a novel administration format has been introduced. In each item the testee has a choice of three rather than one possible classes of response. Firstly he may indicate that a given rule applies equally to all n exemplars in that item in which case it is the best rule. If there is no rule that fits all n exemplars he may indicate the best available rule that fits n-1 exemplars by pointing out the exemplar which constitutes the exception to the rule. If he finds that none of the available rules fulfills either of these criteria, he should indicate that the most applicable rule (s) in that item apply to only $< n-1$ exemplars. In other words, none of the given rules applies to at least n-1 of the symbol groups. In an item containing n exemplars there will be $n + 2$ possible responses open to the subject. The situation which now confronts the testee is a close parallel to the real conditions under which determinative induction as defined, usually operates.

The second broad procedural modification will be dealt with in the following section as it relates directly to the third criterion specification for the test format.

iii) The design of the test should be such that separate dimensions of speed and accuracy are susceptible to independent measurement.

The definition of determinative induction proposes explicitly that speed and accuracy are both relevant parameters in this type of reasoning. The test format should therefore incorporate separate facilities for their independent measurement.

To achieve this the test will consist of a greater number of items than the best subject can reasonably complete accurately within a specified period of time. On a separate answer sheet provision will be made for the recording of responses to each item. The subject will be required to indicate his responses by erasing a blackened strip over the alternative of his choice. If it is the wrong alternative this will be indicated by the appearance of the word FALSE under the blackened area which has been erased. If the subject's response is correct the word TRUE will appear. Subjects will thus get immediate feedback on their responses and will not be permitted to proceed to a subsequent item until the immediately preceeding one has been answered correctly.

The subject's accuracy score will then be a function of the total number of items answered and the number of attempts made in each item. His speed score will be reflected by the actual number of items attempted within the prescribed time limit for the test.

iv) The universe of rules employed in the test items should be clearly and unambiguously formulated and should be provided explicitly in the test instructions. Sample items demonstrating the application of these rules should be provided.

From the point of view of information theory science may be regarded as a useful means of reducing the apparent entropy in the natural world. This is achieved by identifying the central concepts underlying the subject matter of a science and by discerning the general principles governing the way in which these concepts relate to one another. Scientific endeavour toward these goals is guided by a central criterion - the criterion of parsimony. The result is that even the most complex and advanced of todays modern sciences has at its foundations a few kernel concepts and a limited number of general laws. Consider for example the generality of concepts such as distance, time and mass in physics, or the generality of the laws underlying modern electronics.

The complexity of the subject matter of a science is not a function of the number of concepts involved but of their generality. Scientific induction is hence not usually aimed at the unearthing of new concepts, but at the further reduction of entropy by the discovery of rules of greater generality. This is the type of behaviour defined as determinative induction.

By analogy to the situation found in science it is considered therefore, that if a limited number of concepts and a limited number of generally stated laws are incorporated in the test construction model, there should be sufficient scope for the manifestation of a great variety of specific instances of the operation of these laws or rules in the test items. Determinative induction could then be measured in a situation closely paralleling the situation in science.

To fulfill the specifications of our definition it remains only to delineate the concepts used in the test and to formulate explicitly the general rules or broad principles governing them.

A content analysis of a wide variety of inductive reasoning tests revealed the presence of the following common concepts: class membership, replication and series. These basic concepts are cleverly embedded in various guises in the divergent item formats of most tests of induction. It is not difficult to see their relevance to the reasoning required in the real world of science. Furthermore, they are concepts of sufficient generality to permit numerous different manifestations. They are therefore well suited to the requirements of our test model.

The following generally stated rules based on these three concepts have been formulated.

i) <u>Element in fixed position</u>

This rule depends on the concept of class membership. In the general case this rule will be present in an item when any element of a symbol group recurs repeatedly in the same position in all other symbol groups. An element of a symbol group refers to any individual symbol, or to any class of symbols, or to any <u>pair</u> of symbols or to any <u>pair</u> of symbol classes. By the use of the term <u>pair</u> it should be understood that the two individual symbols or symbol classes comprising the pair can appear in any serial order within each pair.

This rule can be expressed more concisely in set theoretic notation:

$$\exists \Omega \forall G (\exists z \vee \exists \{x,y\} \left[P(z,G,_i) \vee Ra_{(i,j)} : (P(x,G,_i) . P(y,G,_j) \vee P(x,G,_j) . P(y,G,_i)) \right]$$

$$Ra_{i,j} \quad |i - j| = 1.$$

where: Ω is the universe of symbol groups comprising an item

G is any symbol group

x,y, and z are any individual symbols or classes of symbols

ii) Duplication of Element

This rule relates to the concept of replication. The general case of the rule will be expressed in an item when an element of each symbol group appears at least twice in the same symbol group. The same meaning as above adheres to the use of the term element.

Again more precision can be given to this rule by expressing it in set theoretic notation:

$$\exists\Omega\forall G.\exists x\exists y \left[G\in\Omega . x\in\Omega . y\in\Omega \rightarrow x = y \right]$$

where: Ω is the universe of symbol groups comprising an item

G is any symbol group

x and y are any individual symbols or classes of symbols

iii) Symbol Series

This rule is based on the general concept of series. For the purpose of this test the rule of symbol series will be expressed in an item when any set of at least three symbols or symbol classes recurs in the same serial order in each symbol group. It is not necessary that the components of the series are always adjacent to one another.

$$\exists\Omega\forall G\exists\{x_i\}_{i=1,2\cdots n} \quad (3\leq n\leq 5)\forall x_i \left[G\in\Omega . x_i\in G.(i<j)\rightarrow R(x_i,x_j) \right]$$

R (a,b) $\equiv$ a is somewhere to the left of b.

where: Ω is the universe of symbol groups comprising an item.

G is any symbol group

x is any individual symbol or class of symbols

These three general rules will be clearly formulated in the instructions to the test. A liberal number of practice items showing their manifestation in specific instances as they may occur in the actual test items will be provided.

REFERENCES

		Order of Appearance
1.	BARKER, J.D. (1969) Handleiding vir Patroonverhoudingstoets (A/15/1): WNNR-gids K7.24, Johannesburg	7
2.	ENGLISH, H.E. and A.C. ENGLISH. (1958) A Comprehensive Dictionary of psychological and psychometric terms. New York, Longmans, Green and Co. inc.	9
3.	FRENCH, G.W., R.B. EKSTROM and L.A. PRICE (1963) Manual for Kit of Reference tests for cognitive factors. New Jersey, Educational Testing Service	2,13,14
4.	GUILFORD, J.P. (1967) The Nature of Human Intelligence. McGraw-Hill Book Co. inc. New York	3
5.	GUTTMAN, L. (1965) The Structure of Interrelations among Intelligence Tests. In: Proceedings of the 1964 International Conference on Testing Problems. Princeton, New Jersey: Educational Testing Service	17
6.	KELLOG, R. (1955) What children scribble and why. Author's edition, San Fransisco	18
7.	KNEEBONE, G.T. (1963) Mathematical logic and the foundations of Mathematics. Van Nostrand, London	11,12
8.	LITTLE, W., H.W. FOWLER and G. COULSON (1964) Shorter Oxford Dictionary. Clarendon Press, Oxford	10
9.	SCHEPERS, J.M. (1970) Die Konstruksie van 'n toets van Konsepverwerwing: Referaat gelewer tydens die Negende Kongres en Jaarvergadering van SIRSA, 29 September 1970. Pretoria	4

10. STEYN, D.W. (1971) 'n Kritiese Waardering van die NIPN toets A/15/1 (Patroonverhoudings) met spesiale verwysing na die konstrukgeldigheid van die instrument. WNNR Verslag Pers. 149, Johannesburg 8

11. THURSTONE, L.L. (1938) Primary Mental Abilities. Psychometric Monographs No. 1 15

12. THURSTONE, L.L. and THURSTONE, T.G. (1941) Factorial Studies of Intelligence. Psychometric Monographs No. 2 1

13. VERSTER, J.M. (1970) Theoretical Model for the Construction of a Test of Deductive Reasoning. Part One: Theoretical Considerations. N.I.P.R., Johannesburg 6

14. VERSTER, J.M. (1972) Deductive Reasoning Test (A/112) N.I.P.R. Johannesburg 5