

CHAPTER



ADMINISTRATION OF THE FINAL TEST AND STATISTICAL ANALYSIS

Once the final form of the test is available, a study needs to be conducted to assess the nature, validity and reliability of scores hypothesised for different factors and to establish norms. The tests and the criterion measures employed, scoring hypotheses and procedure of administration and collection of the data have been described in this chapter, while validity and reliability procedures will be described in the next chapter.

Description of Measures

8.1 Final Test Battery and hypothesised factor score:

Final form of the test battery under study has been described in the previous chapter. A copy of the test battery (CRM I - IV) is given in the appendix C.

A set of heterogeneous measures constituting a battery of tests having relatively high correlation with criterion measures may be more useful and efficient for a practical purpose than would homogeneous set. With heterogeneous measures, the purpose often, is to maximise correlation with outside criterion. Selecting measures in the order of their correlation with the criterion accomplishes the objective (Dubois, 60). Main objective was to see whether different scores for creativity - factors from the same sets of responses using some established ways of multiple scoring contributed significantly to creativity criterion. Such scores which are relatively independent, and contributing significantly to the criterion, combined together would give a composite index of creative ability. Other important objectives would be to study validity and reliability of the composite index to be evolved and to study the nature of individual scores through factor analysis which would lead to some speculation regarding the identification of factors and substantiate some findings of earlier studies.

Hypothesised factor scores: Much has been said earlier about fluency factor. Fluency was originally hypothesised as the ability of creative person to "call up a relatively large number of ideas per unit time" (Guilford, Wilson, Christensen and Lewis: 110). Calling up ideas irrespective of quality

involved ideational fluency. Usually number of responses given to a situation has been taken as IF score. For 'Brick Uses' the number of uses given by S became the fluency score. In Plot Titles and consequences tests, however, only nonclever and non-remote responses respectively formed ideational fluency scores. (Guilford, Wilson and Christensen: 111). In later studies this factor was identified as DTU (Divergent Production of Semantic Units). In circles Test, a score for fluency is simply the number of responses minus the number of duplications and irrelevant responses (Torrance: 262).

Parallel to ideational fluency factor, another fluency factor in the figural content area has been identified. Figural fluency referred to the ability to give 'units of figural information' in unit time. Ss would produce units of figures using the given figures or parts (Ex: Make-a-figure Test) (Gershon, Guilford and Merrifield: 93). Other well known fluency factors are word fluency, associational fluency and expressional fluency. Definitions of these factors have been given in Chapter 3. Each is a divergent thinking ability. Only factors for which scores will be hypothesised will be discussed here. Number of responses given to figures in CRM I and CRM II seem to be definitely fluency scores. Whether each of these scores (from CRM I and II) represent purely single

content categories is a matter for speculation. Considering the nature of operations the scores may either represent ideational or figural fluency factors.

Utility was scored for fluency and flexibility (spontaneous). Total number of responses gave fluency score. Number of classes of responses gave a flexibility score. (Guilford et al: 110). Spontaneous flexibility seemed to present the ability to produce diversity of ideas. "While it requires the individual to change set, it differs from adaptive flexibility, in that the direction of set change is not restricted. The individual is not required to arrive at one particular answer in order to be successful. It might be characterised as lability of ideas" (Guilford et al: 111), Semantic spontaneous flexibility has been recognised as Divergent Production of Semantic Classes (DMC).

In circles Test a flexibility score has been obtained by counting the number of different categories into which subjects responses can be classified. The categories were derived from an analysis and classification of responses of a sample of 588 Ss from Kindergarten through the college years. (Torrance: 262).

In the manner described above responses to CRM I and II - figures can be classified and flexibility score can be obtained.

Not much is to be said about originality as the nature of the score has been discussed often. Out of responses, meaningless, and stereotyped though meaningful responses should be cut off. In other words, number of rare meaningful responses or clever responses will give originality score. That responses can be considered as differentiating original persons from those not original, among the responses to ambiguous figural stimuli has been discussed earlier with ~~xx~~ reference to studies conducted using HIT. Originality seems to be synonymous with uncommonality (Guilford et al. 117). Guilford's way of scoring for originality has been described earlier. The factor originality has been given a logical place in the SI model. The cell is referred by trigram DMT measuring divergent production of semantic transformations. It has been defined as "the ability to produce unusual, remote, or clever responses involving reinterpretation or new emphasis or some aspect of an object or situation." It has been even termed as Semantic Adaptive Flexibility.

In circles test, originality score was obtained by assigning scale values on the basis of statistical infrequency and / or obviousness (Torrance: 262). Wallach and Kogan obtained what they called 'uniqueness score' from their tests by selecting those responses which occurred only once in responses given by 150 children. Responses to CRM I and II

stimulus figures can be scored for originality in the same manner. Hence it is decided that (i) stereotyped, common, oft-occurring, part and meaningless responses should be discounted first (ii) from the remaining responses, all those considered as suitable, clever, beautiful indirect ones showing literary skill should be counted to give originality score.

Elaboration has been identified as divergent production of semantic implications (DMI). It has been defined as production of a number of antecedents, concurrent, or consequences of a given information. Both utility Test (fluency) and Plot Titles (low) which are fluency measures have been found to be loaded with DMI (Guilford et al; 114). Guilford group seems not to have tried the method of multiple scoring for elaboration. Torrance, (262), counted number of pertinent details added to the original stimulus figure itself to get an elaboration score. Hence in CRM I and II, total number of additional ideas or details attached to central idea of each of the responses would form elaboration score. The two scores derived from CRM I and II should be treated separately.

'Fluency in the production of symbolic relationships' was identified in a study done by Hoepfner et al (142.).

Alternate additions in which S is to choose different sets of members from a given set every time to make the same specified total, stood for the said factor. CRM III has been formed after a study of this test. Use of coded information and that number of constituents should be three (triplets) are added restrictions. Total number of responses or triplets completed may stand as a score for fluency of symbols relationships. In S-I terminology, the factor has been named as 'divergent production of symbolic relationships (DSR)'. .

CRM IV may be another approach to the same score. Given a set of five numbers and asked to use all the five numbers and four operations, (+, -, x and \div) only once in a specified manner to get as many positive whole numbers as possible as end products, seems to involve either symbolic relations or symbolic elaboration. Symbolic relations is involved in choosing a number with a relation foreseeing the product to be a positive whole number, just as in alternate additions. Hence number of products got has been taken to be the score.

Manipulation of symbolic relations was the theme of symbol elaboration test. (Hoepfner & Guilford: 142). Given pairs of simple algebraic equations, writing new equations from the given ones was the required operation. Probably

this factor is also involved in CRM IV as the Ss are continuously engaged in different number equations.

Table 9 summarises the scoring hypotheses.

8.2 Criterion Measures:

Glimpses of criterion measures used in some previous studies indicate creativity of individuals can well be rated by teachers, supervisors or laboratory chiefs. Taylor and Ellison (239) found that supervisory ratings of creativity, over-all performance by laboratory chiefs and originality in the written work as among the best of the criteria in preparing a biographic inventory. Ss were air force scientists. For validating a biographic inventory for finding out correlates (biographic) of artistic and literary creativity in adolescent girls, Anastasi and Shaefer (6) used teacher's nominations supported by creative products as criterion. For inclusion in a creative group, S had to meet two criteria: (a) teacher nomination (b) score above a minimum cut off on Guilford Alternate-uses and consequences tests. Sprecher T.R., (231) found that creativity measures similar to those used by Guilford significantly predicted criterion ratings of creativity made by examiners, peers and supervisors.

TABLE 9

Hypothesised Factors and Scoring Procedures for CRM I - IV.

Tests	Code	Hypothesised factors	Scoring Procedures
CRM I	A	Fluency	Total number of relevant responses.
	B	Flexibility	Total number of classes of responses.
	C	Originality	Total number of rare and clever responses.
	D	Elaboration	Total number of additional ideas or details attached to central idea of each response.
CRM II	E	Fluency	Same as above.
	F	Flexibility	
	G	Originality	
	H	Elaboration	
CRM III	I	Fluency	Total number of products completed correctly.
CRM IV	J	Fluency	Total number of products completed correctly.
	K	Elaboration	Total number of sets of operations. (+, -, \div , x)

Maier and Janzen (199) found that supervisor female problem solvers also generate solutions that are rated as creative when several solutions to a given problem are possible. All the studies mentioned above indicate that Teacher's Ratings of pupils' creativity may stand as a 'valid' criterion.

Getzels and Jackson's findings that creative children were less liked as pupils by teachers has been supported by findings by others (Wallach and Kogan : 274, Hasen and Butcher: 134). This makes any researcher to be cautious in accepting teachers ratings as sole criterion. However rating creative performance is different from 'liking or disliking' further if the teachers are furnished with some bases (say a few questions) which they should consider while rating, the effect of like-dislike dilemma may be reduced.

Whether individual factors should be rated separately has been considered in the light of the findings by 'Goodman et al., (105). To see whether each factor score significantly contributes to a wholesome criterion (rated creativity) seems to be better than finding whether each correlates significantly with separate ratings for factors. What human sense can identify or rate, at the most seems to be 'creativity' in performance as it is understood and perceived by the raters and the society to which they

belong, that too at the time of rating. Criterion measures, thus, may be a combination of teacher ratings with some product measure as done by Anastasi and Schaefer (6).

In all three criterion measures were employed in the final study - They are :-

1. Teacher ratings (using a rating scale specially evolved for the purpose).
2. Torrance's Circles Test.
3. Activities Checklist (Prepared on Things - done model).

The foregoing paragraphs give a description of each of the measures.

1. Teacher Ratings: Yamamoto (293) asserts that a reliable criterion measure is desirable but 'its relevancy for the particular purposes of the investigation cannot be determined except by a careful logical analysis.'

It is true, in preparing the Teacher - Rating Scale some logical analysis of operations involved in tests of creativity came to help. The cumulated points creativity rating scale (CPCRS), if it can be called so (Guilford: 113), consists of four sections one for each of the four subjects as viz., Art, Science, Literature and Mathematics.

Each section contains five questions about the S⁴ to be answered by the subject teacher as 'yes' or 'no' by putting '✓' or 'X' mark accordingly against the name of the S. A '✓' mark is equivalent to a score of '1' and in all an S can carry a total score of '20', if the answer by the Teacher is '✓' to all five questions in all four areas.

Question 1. is concerned with S's ability to suggest improvement or give ideas. Torrance's 'Toy Dog' test has been built upon this principle (Torrance, 256). Hence it has been felt, if in teacher's experience the S under consideration has been resourceful in his class room activities or outside, he is creative.

Question 2 is concerned with Ss ability to ask questions which are pertinent to a situation. Torrance's pertinent questions (256) is based on this principle.

Question 3 is concerned with Ss ability to prepare toys, playmaterials etc., This has been utilised in one of Torrance's Tests.

Question 4. is concerned with Ss ability to collect things of interest in that subject area and organising them. Such an interest, it has often been considered as a

creative disposition. Such items have found place in Torrance's things done checklist (256).

Question 5. is concerned with Ss interest in one or the other subject areas. It was felt that such an interest showed the dominant creative disposition.

For example, five questions from 'science' section are as follows:

1. Does he give ideas of his own when you are teaching or conducting experiments?
2. Does he put questions which you have appreciated often while you are teaching?
3. Does he make or try to do scientific toys which he might have seen or read?
4. Does he show interest in making collection of things important to science?
5. Does he show relatively more interest in science than in other class-room subjects?

In other three sections relating to art, literature and mathematics questions were of parallel nature.

If we look into factorial nature of these questions, it becomes obvious to us that ability to give ideas, ask questions, prepare toys, collect things of interest involves such factors as ideational fluency, sensitivity to problem: originality, flexibility etc., They also seek

long range interests and persistent efforts. Last question, however, tries to trace the dominant trend of US creativity. All these points lead the author to believe that a total score would be definitely a good criterion measure of creativity.

2. Torrance's Circles Test (256): The circles test is one of the three nonverbal tests of Minnesota Tests of Creative Thinking. Brief description of this test has been given in Chapter IV. The test has been taken as one of the criterion measures. However only one test out of the three tests in the nonverbal form has been chosen and is a departure from the usual practice. This has been done on the following grounds:

(a) The operations involved in the test seemed similar to the operations involved in CRM I and II (This point has been discussed while giving a rationale for the test). As the circles are repeated as stimuli, it is felt that it is as good as giving single stimulus and asking them to write what different things they would make out of it. For example see Wallach and Hogen's 'instances' eg., "Name all the round things you can think of".

(b) The test gives scores for fluency, flexibility, originality and elaboration simultaneously derived from

single set of responses. As deriving multiple scores and retaining those which are relatively independent among themselves and predictive of the criterion would be the strategy, it would be convenient to study the scores similarly derived in CRM I and II.

(c) The circles test would be the sole representative of figural stimulus - figural response type tests. As Torrance has considered, ideational fluency does play a part. It would be interesting to see how ideational fluency score on verbally expressed ideas will be related to ideas expressed figurally.

(d) The other two tests of DTCT (or Minnesota) non-verbal form are operationally different from 'circles' test. Hence including all the three would give impure measures.

Keeping the above points in mind the investigator preferred to choose 'circles test' only, to be included in the set of tests to be administered in the final study. However including a few tests from Guilford's group would have been useful. As described in the previous chapter, there is no single test which would give all four scores as 'circles'. The scores have been derived by Torrance on the basis of definition given by Guilford. The concepts

of ideational fluency, spontaneous flexibility, originality and Elaboration have respectively stood for number of ideas, number of classes of ideas, cleverness or statistical rarity and number of details in either's case.

3. Activities Checklist: This is more or less analogous to Torrance's 'things done on your own'. (256). Ss are to check the activities they did on their own in the course of one year. Some of the items in Torrance's list were unsuitable to Indian conditions. These were removed and some activities were added. Total number of activities remained however same at 100. In addition list was made open-ended by providing space for writing activities added by the S. Another difference thought out was to get score based on statistical rarity of activities rather than total number of activities checked by the S. Procedure adopted will be described in the appropriate section on 'scoring'.

8.3 Collection of the Data:

Sample: Sample in the final study consisted of 230

X std., children from four schools of Udipi Taluk (Mangalore Dt.), Mysore State.

Co-operation of the headmasters and the staff in all the four schools was excellent. In all 231 (boys and girls) appeared for the test. Only one boy was to be discounted for the reason that he gave one of the papers blank. Testing was done in groups of 40-50 children. However this was not a strict rule. Usually a class which varied in its strength was taken a group of study.

Administration: Final Tests (CRM I - IV) along with two criterion measures viz., (1) Torrance Circles Test and (2) Activities checklist were administered to children in the same fashion described in the pilot study. The only difference was that the final tests were short. Hence testing took two hours (approximately) including the time of instructions and was done in single sessions. The investigator himself conducted the testing. Instructions were orally given as described earlier (see Chapter 6). As usual each item was individually timed, for Torrance's Test, directions given in the manual (Torrance 262) were followed. Ss were supplied with activities checklist and the investigator read each of the activities one after the other while the Ss checked those activities they had done. Sufficient time was allowed for writing about activities they had done but not found in the list. Order in which the tests were administered was as follows:

CRM I, II, III, IV, circles and Activities Checklist.

Teacher ratings were taken separately from teachers nominated by the headmaster. The investigator acquainted them with the purpose and the rating scale they had to use. They were given 2-3 day's time to complete the rating. The rating scale (CPCR3) has been described earlier.

Scoring: Scoring keys were necessary for flexibility and originality scoring in the case of CRM I and II, and for number of correct products in the case of CRM III. Flexibility and originality keys were developed on the basis of responses given by 100 Ss from the pilot study sample. The 100 Ss (50 from each of the two areas viz., Udipi and Dharwar) were selected randomly. Each of the responses were listed itemwise. Classes were decided arbitrarily. The investigator's discretion combined with similarity of response attributes lead the classification. Keys for originality scoring were developed as follows: Those responses which occurred four times or less were retained as rare responses. The lists of responses (itemwise) were supplied to two judges asking them to decide which of the responses were obviously not original. Responses for which the decision of the two judges concurred as not original were rejected. These keys however got

enlarged while scoring, by addition of few more responses which did not occur previously and were considered to be original by the investigator. This amount of freedom has been made possible in other tests too (Torrance: 262, Barron: 23). The keys for originality scoring have been given in the appendix D.

Circles test was scored according to the keys given in the directions and scoring manual developed by Torrance (262). Activities checklist was scored in the following manner - Checklists from ninety Ss were examined for frequency of checks for each of the 100 items. Those items which were checked by more than twentyfive Ss were rejected. Thus fortysix items were to be rejected. Only fiftyfour items were regarded for scoring. Number of items checked out of these fiftyfour items plus a weight of three (arbitrarily decided) for each additional activity written by the S at the end formed the Ss checklist score.

In the following section of this chapter statistical analysis of the data has been presented.

8.4 Statistical Analysis of the Data:

In the previous section, how 11 scores from the four tests of the battery CRM I - IV and the six criterion

measures from three different procedures namely Torrance's circles test, Teacher's Ratings and Activities checklist were obtained was described. Of these, the last measure obtained through checklist will be used cautiously at a later stage. The present chapter described the nature of the remaining sixteen scores statistically, thus finalising scoring procedures to be retained to give a composite creativity measure. Addition and multiplication machines were employed in computing all statistics to be given hereafterwards.

Means and S.Ds: Means and standard deviations of the sixteen measures (raw scores) have been presented in Table 10. These have been calculated using the formulas (1) and (16) given in Garret (90). The sixteen measures have been coded using alphabets viz., A to P. Hereafterwards scores will be referred through respective codes. This has been simply for the sake of convenience and codes do not carry any other abbreviated meaning except that they stand for the respective hypothesised factor content.

Transformation into Comparable Units:

Obviously some of the distributions seem to be positively skewed. It is, however, decided to treat the distributions as they are, for the reason that

TABLE 10

Means and Standard Deviations of the 16 measures of
Creativity (CRM I - IV and criterion
measures) for 230 X std. children

Score	Mean (M) (from raw scores)	Standard Deviation (G)
A	19.0304	6.6256
B	11.7826	2.7522
C	3.5478	2.5276
D	22.6261	12.4785
E	17.1826	6.7122
F	13.6304	4.6497
G	3.6652	2.4166
H	12.6913	7.1582
I	6.2043	2.5584
J	3.3913	2.7000
K	2.3826	1.6075
L	6.3088	4.5057
M	4.9261	3.3147
N	4.7044	4.8991
O	11.5000	8.4230
P	4.3870	3.5829

clusters of scores namely (A, B, C, D), (E, F, G, H), (I, J) and (K, L, M, N, O) have been derived from single sets of responses to individual tests namely CRM I, II, IV and Torrance's Circles Test and bear interrelationships among themselves. Any normalisation would tell differently on different score - distributions and hence interrelationships will be affected.

In order to make comparison possible, raw scores have been converted into standard scores of mean of 50 and σ of 10 using linear transformation equation no. 19.2 given by Guilford (113) given in Chapter 5. For all future calculation standard scores would be employed and would be treated as if they are raw scores.

A study of intercorrelations: In order to study the nature of relationship among the eleven creativity measures derived from the tests under consideration and to see how many of these measures are significantly related to criterion measures, coefficients of correlation among all the sixteen measures have been computed. As already pointed standard score equivalents of raw scores have been employed in computing intercorrelations. Formula (32) given by Garret (92) has been utilised for computations. Inter-correlations among the sixteen measures have been presented in Table II.

Nature of each of the eleven measures under study in terms of significant relationships with the five criterion measures as revealed by inter-correlation matrix has been discussed in the following paragraphs.

'A', as hypothesised, has shown significant relationship with fluency score 'B' from circles test. A has no significant relationship with 'P'. There is no sufficient ground to believe that score 'B', hypothesised to go with flexibility score 'H' from Circles Test. For it has any significant relationships with Teacher's Ratings. Instead it has gone with fluency score.

'C' as hypothesised, has significant relationship with 'B' originality score from Circles Test and highly significant relationship with 'P' (teacher ratings) indicating that teachers have regarded 'originality' in their ratings.

'D', hypothesised to go with elaboration has significant relationship with teachers ratings. It is quite possible that teachers have regarded elaborative ability. However D has no significant relationship with elaboration score 'C' of circles test. It is also possible that 'content' of elaboration score differed in the two tests.

TABLE 11

Correlation Matrix: Showing inter-correlations among
16 creativity measures obtained from a
sample of 230 X std. children

	A	B	C	D	E	F	G	H
A		846	499	438	693	698	436	438
B			483	413	-504	562	450	440
C				520	390	411	499	377
D					417	452	379	674
E						919	517	483
F							537	470
G								351
H								
I								
J								
K								
L								
M								
N								
O								
P								

Note: Decimal points omitted.
*For df 228 $r \geq .13$ significant at .05 level.
*For df 228 $r \geq .17$ significant at .01 level.

TABLE 11
Correlation Matrix
(Continued)

	I	J	K	L	M	N	O	P
A	178	007	-063	167*	114	-034	084	016
B	242	127	023	131*	093	-082	069	080
C	054	007	-069	018	040	130	094	241*
D	086	031	-023	-049	-022	071	084	209*
E	072	010	-063	043	045	-028	-045	040
F	083	038	-061	051	066	006	000	032
G	019	005	025	072	096	188*	113	138*
H	-029	097	-021	-024	-032	071	199*	227*
I		411	347	269*	337*	092	179*	293*
J			791	108	211*	-019	096	245*
K				098	-093	-073	019	259*
L					872	469	736	143
M						499	765	132
N							549	159
O								134
P								

Note: Decimal points omitted.

*For df 228 $r \geq .13$ significant at .05 level.

*For df 228 $r \geq .17$ significant at .01 level.

It is notably strange that 'Q' has failed to go with any of the criterion measures. Same is the case with 'F' showing little reason to believe that F differed from 'Q'.

It is important to note that 'G' as hypothesised, has shown significant relationship with 'H' and 'P' as its peer 'C' did.

'M' has shown significant relationship with 'O' (elaboration) and also with teachers ratings (P).

'I' supposed a symbolic test, score shows significant relationship with three of the circles test scores, L, M and O; and also with teacher ratings (P). Even though factorial nature of this score cannot be speculated here, the relationships are sufficient to retain it as a creativity measure.

'J' differs from 'I' in that it has significant relationship with M and P only. K differs from J and I in that it is related significantly with P only.

Lowest of the correlation coefficients among the measures of the first cluster A, B, C, and D is .41 and lowest of the correlation coefficients among the measures of the second cluster E, F, G and H is .35. This is

comparable with the lowest of the correlation coefficients of similar measures from circles test cluster (L, M, N and O), which is .469. This seems to be sufficient to show that the multiple measures derived from CRM I and II differed among themselves as much as multiple measures from circles test differed among themselves.

Measures I, J and K being symbolic in content differ from majority of measures of CRM I and II as evident from not significant correlation coefficients. Secondly they have remarkably positive significant relationships with TCT measures as well as with P.

In order to retain any score in the composite measure to be developed, the following conditions seemed reasonable.

To be retained, a measure should be (i) significantly correlated with at least one of the criterion measures (this is irrespective of factorial orientation which is yet to be studied) (ii) different from its peers in cluster (to which it belongs) in terms of its relationships with criterion measures. These points have been made clear in the above discussion.

In conclusion, it can be said that except B, E and F all measures are qualified to give a composite creativity index. B, E and F scores have been rejected. A, C, D, G, H, I, J and K have been retained to form a composite measure of creativity. This will be hereafterwards referred to as 'composite creativity index' or C.C.I. which is nothing but a simple sum of standard scores of eight variables chosen viz., A, C, D, G, H, I, J and K.

Mean and S.D. of the composite creativity index: As means, s.d.s and intercorrelations for the eight creativity measures are available, mean and s.d of their composite (i.e., composite creativity index) obtained by adding the standard scores, have been computed using the formulas A .33 and A .34 given in Guilford (1959) respectively. Mean and S.D. of the C.C.I (N=230) have been found to be respectively 400.27 and 45.82.

Next chapter deals with validity and reliability of the composite measure besides throwing light into factorial nature of the scores. An attempt to establish norms has been made.