

STRESS REACTIVITY RESEARCH AND PSYCHOPATHIC INSTABILITY

J. E. TONG

Department of Psychology, University of Auckland, New Zealand

Introduction

A recent paper by Esher, Orme and McKerracher (1965) describes a study using techniques derived from earlier work by the present writer and associates at Rampton Hospital. Esher et al consider that

(a) their results do not support previous claims, and
(b) their study was "replicatory" and should have been able to "reduplicate" previous findings.

This present note disagrees with the Esher claims and points out

(a) their results, which appear to involve at least one statistical error, add some support to earlier published material concerning "low stress reactivity" as a possible form of psychopathic instability, and
(b) they do not use all the available published figures for comparison of scores, and

(c) they make an invalid comparison of scores, and
(d) by introducing important variations in technique, and research design, their study is not a true replication, and

(e) they misconstrue the published theoretical basis of the earlier work.

Published material relevant to this discussion is in Tong, 1957; 1959; 1960a; 1962; Tong and Mackay, 1959; Tong and Murphy 1960a, 1960; and Tong, Murphy and Adams, 1960. Relevant unpublished data are available in Murphy, 1961. A paper (Tong, 1960b) quoted by Esher *et al* is irrelevant and deals solely with the sex offender.

G.S.R. Conditioning Scores and Relapse

In the 1959 paper it was stated "The ultimate purpose of this research (on the Rampton patient) is to determine the structural personality differences which contribute towards relapse and continued instability". It was repeatedly argued at that time and will be restated below that the only satisfactory criterion against which one can compare test scores for delinquents and psychopaths is a reasonably objective social criterion, preferably antisocial behaviour on discharge from hospital or institution. Esher et al give no figures relating their scores to such a criterion.

A very large number of laboratory test scores were analysed in relation to each other and to the social criteria by the writer and associates. Of these the most extensive information is on the G.S.R. Conditioning Score, partly because it is from the most widely applicable technique. Such a score is also used by Esher et al.

The published figures for the first follow-up of 45 discharged Rampton patients (Tong, 1959) were as follows:—

	Conditioning Score	
	<7	>6
Success Group	16	10
Relapse Group	18	1

These figures indicated that Relapse patients were almost all of Low score.

A second follow-up study of different patients (Tong, 1962) gave the following results:—

					Conditioning Score		
					0-3	4-8	9-11
Success Group	11	28	9
Relapse Group	19	8	8

It was suggested (1962) "Poor stress reactivity corresponds with instability and moderate stress reactivity with stability, but obviously high stress reactivity is a more complex problem".

Murphy (1961) examined *selected* patients of whom insufficient were discharged to provide adequate follow-up material. Hence his Relapse Group comprised both patients tested before discharge and patients first tested after relapse. These figures are:—

					Conditioning Score		
					0-3	4-7	8-11
Success Group	8	8	16
Relapse Group	10	3	4

These data support the previous suggestions concerning Low score, although they concern specially selected patients only.

On the basis of these social criterion data one would expect that a sample of Rampton patients being extremely socially unstable, by definition, would contain more people of Low Conditioning Score than would a sample from an ordinary open hospital such as Hollow Meadows. Esher et al report such differences at the 0.0005 level of P, in the expected direction. *Hence their data provide independent support for the earlier suggestions* by the present writer concerning the social/clinical significance of a low score on this laboratory measure.

Still further evidence of the predictive significance of the low score is also given by Esher et al by showing that 17 of the 21 low scorers assessed five or more years earlier are still rated unstable by hospital staff. No conceivable testing bias could possibly account for such *predictive* accuracy, and no other *single* measure from clinical tests nor personal assessment could discriminate in such a fashion.

General Distribution of the G.S.R. Conditioning Scores

Esher et al compare a sample of scores for Hollow Meadows, 1964, with a sample for Rampton, 1964 and Murphy's data for Rampton, 1960. They draw clinical inferences from the U shaped distribution of the latter compared with the different distributions for the two former groups. They fail to point out that the latter were aggression cases only. They also conclude from a comparison of the latter two groups that there has been a *significant* demographic change in the Rampton population. *Neither conclusion is valid because the data are not comparable.* Murphy's 1960 figures refer only to highly selected patients and were not considered representative of the total Rampton population. Murphy's subjects were only those clearly defined as *physically aggressive* according to stringent criteria. Other data relating types of disorder to stress scores are in Tong (1959).

The only published figures (not used by Esher et al) for what approaches a representative sample of the earlier Rampton population are in Tong (1962) for 401 patients. The histograms do not indicate a U shaped distribution and are *not* significantly different from the Rampton figures given by Esher for 1964, ($X^2=5.77$). The figures are:—

					<i>Conditioning Score</i>		
					0-3	4-7	8-11
Tong, 1962	142	149	110
Esher, 1965	38	32	13

The inclusion of mentally ill patients by Esher et al further invalidates their comparison. In all reports by the present writer and associates it was emphasised that patients *suspected* of psychotic disorder or brain damage were excluded. Similarly, patients receiving drugs were excluded, but this is not considered by Esher.

Nurses' Rating of Stability

Much of the dissatisfaction shown by Esher et al seems to rest upon a failure to find a relationship between their scores and ratings of stability by nurses when in fact, such a relationship is clearly present. They state "Tong and Murphy had reported that skin temperature decrease and a moderate degree of conditionability were closely related with nurses' rating a patient's behaviour stable rather than unstable". *No such simple general claim has been reported.* From the information given by Esher et al it appears that their rating of stability bears little relationship to that reported by Tong (1957, 1959) or Murphy (1961). The difference is that Esher classifies all of his subjects whereas the other studies were concerned only with the selection of clear cut groups of stable and unstable patients. When subjective impressions (i.e. ratings) are used as the criteria against which to relate test scores this is a fundamental psychometric difference.

For the 1959 report 200 names were each rated by 3 judges as Stable, Unstable or Not Known. Only those names against which at least 2 judges recorded Stable or Unstable were related to test scores. Hence the members of the final groups were well known to the judges who agreed on their stability rating. The stated aim was to obtain two fairly clearcut groups and not to relate the scores to a total hospital population as is implied by Esher et al, and undertaken by them in 1964.

The 1959 figures show that only 67 out of the original 200 met the requirements of this aspect of the study for the Conditioning Score. The published figures are:—

					<i>Conditioning Score</i>		
					0-3	4-6	7-11
Stable Group	15	18	10
Unstable Group	13	2	9

Using a similar rating method, Murphy found *no* significant relationship between Conditioning Score and classification for hospital stability for aggressive patients. He found, however, a slight relationship between another conditioning measure and stability rating. This is discussed below.

Mirror Drawing and Temperature Change

Inconsistencies in Esher's paper make for difficulty in interpretation. They state that earlier work indicated that temperature decrease was closely related to nurses rating a patient as stable rather than unstable, yet quote Murphy's figures for a mirror drawing task, when these show no relationship. *Furthermore they state incorrectly that their own figures show no association between temperature change and stability when there is a significant relationship in their table.*

Using a totally different task on subjects selected on the basis of their ability to complete earlier parts of the task, Tong (1959) reported a slight association between surface temperature change and stability rating as described above. The published figures were:—

	Temperature Change	
	Decrease	Zero or Increase
Stable Group	17	12
Unstable Group	4	13

For the Hollow Meadows patients, if Esher's figures are classified in a similar way, the value for X^2 is 4.05, ($P=0.05$) which indicates that *there is a relationship between the stability rating and the physiological measure, contrary to their assertion.* Classifying zero change and increase in temperature together is usually justified by the argument that temperature decrease is a recognized stress reaction, whereas increase may or may not be a stress reaction owing to the possibility of procedure artifacts. *Esher's figures, therefore, again indicate quite clear independent support for the notion of low stress reactivity in relation to stability.*

It is of interest that although both sets of data (Esher, 1964 and Tong, 1959) show significant relationships with stability ratings the contingency tables are of different structure. Skin temperature indices of stress reactivity are extremely complex, as various extraneous factors affect the scores. Murphy, for example, pointed out the calculated significance of "time of day", and clearly room temperature must be standardized. The clinical significance of scores is by no means clear as was apparent from the 1959 report. It was shown then that normals and patients subjected to similar tasks show dissimilar temperature curves. The decrease for normals was slightly greater than that of patients, and there were great differences in recovery following the application of the stressor. Hence, no simple hypothesis of the type propounded by Esher et al is possible at the present state of knowledge. Nevertheless, the Esher data, when analysed are encouraging evidence for the value of physiological correlates in a general sense.

Reliability of Conditioning Scores

Again the data provided by Esher are difficult to interpret. The correlation between the R Score and the CR Score for their Rampton subjects is similar to those determined earlier which suggests some similarity between the earlier and the current techniques. For an unstated reason they do not give the correlation between R Score and CR Score for Hollow Meadows patients, however, so no comment is possible.

Esher introduces novel statistical terms, stating ". . . . Though the correlation between PR and CR or UCR is consistent it is not nearly so *substantial* as the correlation between CR and UCR. The *closer* association in the Hollow Meadows' patients of PR and CR raises. . . .". If "substantial" and "closer" imply statistical significance, then the Hollow Meadows co-efficient is not significantly different from the Rampton co-efficient.

They report a Pearson correlation co-efficient of $+ .18$ between scores of 1958-60 and those of 1964. This they dismiss as non-significant when it is only just below the 0.05 level for 74 subjects. Thus there may be some long term reliability which could be demonstrated by a more adequate statistical technique. There is no reason why one should expect this measure to be highly reliable. For intervals of between 3 months and 12 months Murphy (1961) determined reliability at the 0.01 level using tau statistic. It was pointed out that although the score seemed to be stable for the majority of subjects, a few swing from one end of the scale to the other (Tong, 1960). These latter patients *appeared* to be grossly unstable. If this is so one would expect them still to be at Rampton and reduce the size of a reliability co-efficient. Esher et al give no indication of the reliability of the recent scores, and although they show the relationship of 1958/60 scores to 1964 stability ratings for the 74 available subjects they do not show their present scores in relation to stability ratings for these 74 subjects. Without these data further comment is pointless.

Standardization of Test Procedure

It has been mentioned that there is little similarity between the rating methods of Esher et al and the original studies. Similar comments have been made in connection with the stressors originally used to produce skin temperature change. However, the Esher report also suggests that they modified the mirror drawing task from that used by Murphy. Esher states "The mirror drawing task consisted of the patient tracing round between the double lines of a five pointed star which could be perceived only in a mirror". The Murphy task required six minutes of mirror drawing activity. plus verbal motivation.

Concerning the G.S.R. conditioning technique, Murphy (1961) used the eye-touch (S) and buzzer (CS) method of the present writer, and also applied a very similar technique on the same day to the same patients aimed to provide improved scores. The second G.S.R. conditioning method used a puff of air to the eye (S) and a light (CS). Although these two methods appear to be alike the scores are not completely isomorphous. A 3×3 contingency table reached the 0.001 level of significance, with a contingency co-efficient of .29. Of 161 subjects tested with both methods 71 were "misplaced" in the table. In the analysis of nurse's ratings undertaken by Murphy the first scores were not related to the criterion, but the second were. The reverse was the case in his follow-up data.

For the research undertaken by Esher et al it has been well established for some years that what appear to be minor variations in technique can influence results in a marked fashion. Hence the modifications introduced by Esher *without first determining the effect on the scores* makes for weak comparison with earlier figures and results. Information from current research in New Zealand would suggest that differences in testing procedure would contribute to the phenomenally large differences in scores between Rampton and Hollow Meadows subjects.

Theory

Esher et al make references to Pavlovian theory and arousal theory. They appear to confuse conditioning and arousal when they ask:— "Is there any point in carrying out conditioning experiments when simpler measures of arousal are available?" Several important points are involved here.

The writer and colleagues have avoided Pavlovian interpretations in all theoretical comments, owing to the lack of precision in that theory. All measures have

been regarded as "stress measures" simply because reaction to stress appears to be a workable hunch for the episodic instability of delinquents and psychopaths. If one wishes to regard the stress measures in Pavlovian terms, as Esher infers, then there is little concordance between conditioning theories and theories of arousal or activation. If on the other hand one attempts to use an arousal theory framework, one would anticipate very little relationship between impressionistic assessments of stability and laboratory measures. Arousal theory distinguishes sharply between the *direction* of activation and the *intensity* of activation. These two dimensions are almost bound to be confused by "personal impression" methods of assessment. The vast area of research relating physiological measures and behaviour is well reviewed by Duffy (1962).

The conditioned G.S.R. was chosen in preference to other G.S.R. measures because of its administrative simplicity and the relative freedom from artifact interference. The drive function of stress or anxiety must be considered for any theory of delinquency using stress variables. Consequently, a conditioned or learned response is more directly relevant than a simple stress response. Frequently, much of the clinical problem of the psychopath and the delinquent is that the instability is purely social, episodic, and even quite specific. Diagnosis is based almost entirely on the social history, there being no tangible clinical features for the psychiatrist or nurse to identify and treat. In the early Rampton research, the *general* correlates of instability were first isolated in a large scale follow-up (Tong and Mackay, 1959). The stress research studies were aimed to isolate laboratory correlates of *social* instability on the assumption that any other criterion of instability would be too weak for conclusive experimentation. In 1960 the writer commented, . . . "Ultimately the test by which laboratory data will stand or fall is the degree of demonstrable relationship with real life criteria, and in particular the extent to which they enable one to predict future events accurately and devise therapies".

Conclusion

Esher, Orme and McKerracher (op.cit.), conclude " . . . the results of Tong and Murphy are not generalisable outside the particular groups and conditions from which they were obtained". The foregoing analysis indicates quite clearly that the converse is the case in so far as the Esher techniques are comparable. When adequate data are used for comparison purposes their figures show no significant differences between past and present Rampton patients for the conditioning score. Their figures show a high predictive validity for the low score, and the difference between Rampton and Hollow Meadows patients is in the expected direction. Correct computation of their published figures for temperature change show significant trends in the direction suggested by earlier research. Making allowances for the differences in technique, the data by Esher et al, seem to be important independent clear corroboration of earlier suggestions concerning the feasibility of laboratory correlates for social instability.

The Esher report highlights the necessity of using adequate criteria of stability in research with psychopathic subjects. Impressionistic ratings of stability are notoriously weak and are meaningful only when safeguards are introduced. One such safeguard used in the early Rampton research was to ensure high inter-judge reliability. Esher et al report no such safeguard and their figures indicate that "stability" has a *different meaning*, for example, for Hollow Meadows staff than Rampton staff. This leads to the odd information in their table that significantly more patients in the open hospital are unstable than is the case for the Special hospital.

References

- Esher, F. J. S., Orme, J. E. and McKerracher, D. W., (1965). Replicatory Studies of two psycho-physiological techniques of assessing mentally subnormal patients. *J. Ment. Subnormal.* 9, 2.
- Duffy, E., (1962). *Activation and Behaviour*. Wiley. New York.
- Murphy, I. C., (1961). *Stress reactivity and antisocial aggression*. Ph.D. Thesis, Sheffield Univ. Library.
- Tong, J. E., (1957). Assessment of anxiety as an intervening variable. *Brit. J. Psychol.*, 48, 1, 12.
- Tong, J. E., (1959). Stress reactivity in relation to delinquent and psychopathic behaviour. *J. Ment. Sci.* 104, 441, 935.
- Tong, J. E., (1960a). Psycho-physiological differences in conduct disorders. *J. Ment. Subnormal.* 6, 2, 11.
- Tong, J. E., (1960b). Galvanic skin response studies of sex responsiveness in sex offenders and others. *J. Ment. Sci.*, 106, 445, 1475.
- Tong, J. E., (1962). Psychophysiological studies in psychopathy and the prediction of stability. *Proceedings of the London Conference on the Scientific Study of Mental Deficiency*. Vol. 1, p. 97. Ed. B. W. Richards, May and Baker, Dagenham.
- Tong, J. E. and Mackay, G. W., (1959). A statistical follow-up of mental defectives of dangerous or violent propensities. *Brit. J. Delinq.*, 9, 4, 276.
- Tong, J. E. and Murphy, I. C., (1960a). A review of stress reactivity research in relation to psycho-pathology. *J. Ment. Sci.*, 106, 445, 1272.
- Tong, J. E. and Murphy, I. C., (1960b). Rorschach indices and autonomic stress reactivity. *J. Clin. Psychol.* 16, 3, 324.
- Tong, J. E., Murphy, I. C. and Adams, F., (1960). Preliminary psychomotor stress studies with subnormal psychopathic subjects. *Nature.* 187, 4732.