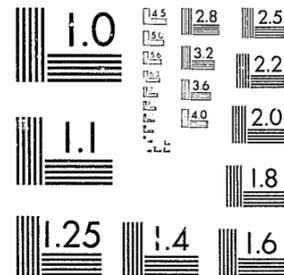


National Criminal Justice Reference Service



This microfiche was produced from documents received for inclusion in the NCJRS data base. Since NCJRS cannot exercise control over the physical condition of the documents submitted, the individual frame quality will vary. The resolution chart on this frame may be used to evaluate the document quality.



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Microfilming procedures used to create this fiche comply with the standards set forth in 41CFR 101-11.504.

Points of view or opinions stated in this document are those of the author(s) and do not represent the official position or policies of the U. S. Department of Justice.

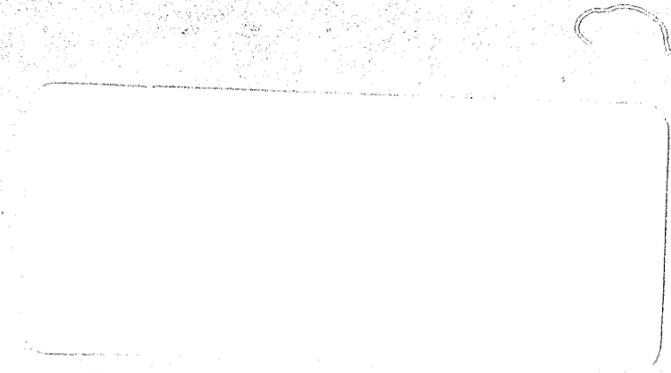
National Institute of Justice
United States Department of Justice
Washington, D. C. 20531

6-17-82

ME1



WORKING PAPERS IN FORENSIC PSYCHIATRY



METROPOLITAN TORONTO

FORENSIC SERVICE

METFORS

81785

U.S. Department of Justice
National Institute of Justice

This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated in this document are those of the authors and do not necessarily represent the official position or policies of the National Institute of Justice.

Permission to reproduce this copyrighted material has been granted by

Suzanne Neuman

METFORS

to the National Criminal Justice Reference Service (NCJRS).

Further reproduction outside of the NCJRS system requires permission of the copyright owner.

ASSESSING DANGEROUS BEHAVIOUR BY MEANS OF
VIDEOTAPED INTERVIEWS:
DATA ON INTER-RATER RELIABILITY BASED ON
A NEW COMPREHENSIVE SCALE

D. Slomen, C.D. Webster, J. Dacre, D. Sepejak,
B.T. Butler, F.A.S. Jensen and G. Turral

W.P. #18

This manuscript comprises unpublished material, not in final form suitable for publication. Please do not quote from it or use it in any way without the written permission of the authors.

This project was supported by Ontario Ministry of Health Grant DM395.

A version of this paper was presented at the 1980 Meeting of the Ontario Psychiatric Association, Toronto, January 24 - 26, 1980.

METFORS
999 Queen Street West
TORONTO, Ontario
M6J 1H4

NCJRS

MAY 23 1980

ACQUISITION

Assessing Dangerous Behaviour by Means of
Videotaped Interviews: Data on Inter-Rater
Reliability Based on a New Comprehensive Scale

D. Slomen, C.D. Webster, J. Dacre, D. Sepejak, B.T. Butler,
F.A.S. Jensen and G. Turrall. Metropolitan Toronto Forensic Service

Elsewhere we have suggested why it is difficult to design and execute studies aimed at determining the extent to which psychiatrists and other mental health workers can predict dangerous behaviour of mentally disordered offenders (Webster, Butler, Jensen and Turrall, 1978, unpublished; Sepejak, 1979, unpublished). And in a related report we have described in considerable detail how we have with clinicians developed a scale for the assessment of dangerous behaviour (Slomen, Webster, Butler et al, 1978, unpublished). In that article we define the 23 items which constitute our scale and describe how 200 subjects were rated by members of interdisciplinary teams and by a pair of external non-involved clinicians.

We cannot here review the entire main project of which the present study is a minor part. Yet, it is necessary for the reader to know that the results reported here depend directly on the outcome of the previous main study. More specifically, we were able in that study to demonstrate acceptable inter-coder agreement on 17 of the 23 items which constitute the scale.[1] The present sub-project is founded

1. Based on intra-class correlations using the two external coders who evaluated all 200 patients. A report on this project is in preparation.

upon the notion that, in all probability, experienced forensic clinicians came to form judgements on the basis of quite subtle cues. Such cues arise during the course of interviewing and can best be studied by means of videotaped records of interviews between patient and forensic clinician. We have explained the rationale in some detail elsewhere (Webster, Slomen, Butler, et al, 1978, unpublished).

The present study was designed to find out: (1), if our previous findings of acceptable inter-rater reliability with the Dangerous Behaviour Rating Scheme (DBRS) could be confirmed with different coders; (2), if the opinions of coders formed through the analysis of videotaped records accords with judgements formed by the interviewing psychiatrist; and (3), if coders are capable of forming judgements even when the amount of information available to them is sharply limited (to the audio channel or to the visual channel).

Before proceeding to describe further the design of the present project it is necessary to explain that the Metropolitan Toronto Forensic Service (METFORS) is a specialized forensic unit which provides assessments for the Courts in the city. Outpatient assessments are completed in one day within the Brief Assessment Unit (BAU). Others are more protracted and take place within the 23 bed Inpatient Unit. Most assessments are completed at a pre-trial level. Normally these assessments do not last more than 30 days. In the present study we examined nine male inpatients. Ages ranged from 19 to 51 with a mean of 30.2 years. Patients were considered for

inclusion if they capable of giving informed consent to the televised interview, if they were male, and if they could be interviewed reasonably soon after admission (in order that the tape could be erased before departure). All interviews were conducted by the same thoroughly experienced forensic psychiatrist and all were approximately 40 minutes long. Although the examining psychiatrist did not see the patient's file before the interview, he was free, during the course of conversation, to inquire about the current offence, psychiatric history and general family and social background. During the interview the patient and psychiatrist sat close together in chairs placed at a 45 degree angle. The recording equipment and camera operator were in full view. After the session was complete the patient and psychiatrist each completed the Session Perception Questionnaire. When the patient left to return to the ward, the psychiatrist completed the DBRS.

We can now consider the design of the study. In Figure 1 we see that there were three pairs of coders^[1] and three viewing conditions: A + V - Audio plus Visual (i.e., the standard arrangement); Ao - Audio Only (i.e., video turned off); Vo - Video Only (audio track turned off). Ratings were made four times for each subject, at the end of each block. Length of block was determined simply by dividing the total interview time by four. It should be evident from Figure 1 that

1. All coders held at least a B.A. degree in the social sciences. In fact, four held M.A.'s and one was close to completing the M.A. Three of the four M.A.'s were, at the time the study was in effect, proceeding to Ph.D.'s. Four coders were female and two were male.

in the fourth block all coders were treated the same way. That is, audio and visual information was available. It will also be apparent that, though beyond the scope of this particular presentation, the design of the study is such that we can to some extent determine the relative influence of audio and visual information. We are particularly interested to learn the degree to which raters can form judgements about dangerousness given non-verbal cues only. This can be accomplished by examining the coder pairs across the first three blocks. We are, however, not concerned with this question at the present time.

Raters were not forced to make a judgement on each item at the end of each block. If they so desired they could tick a column for "Do Not Know" or "Not Applicable". Pooling these two categories we see from Figure 2 that, even after the end of the first block (i.e., after some 10 minutes of viewing, listening or viewing and listening), they were able to offer an appreciable number of definite opinions. We also see that the curves rise steadily over blocks. Figure 2 simply shows that some judgement was being stated. It tells us nothing about the reliability of that opinion.

To gain an impression of reliability we must turn to Figure 3 where we see data from the three coder pairs and, at the bottom, the interviewer psychiatrist. It is important to recognize that these histograms are based on ratings made at the end of the fourth block

(i.e., after the tape had run its full course and after all coders had had at least one block of exposure to both audio and visual stimuli.)

A word must be said about the "composite dangerousness rating score" plotted on the ordinate. The score was obtained by pooling the items found in the previously mentioned study to yield acceptable inter-judge agreement. Excluded were items found unreliable and items, though reliable, either global (eg., "dangerous to self at present") or unrelated to dangerousness per se (eg., "Is Individual Manipulative? Did Individual Provide Accurate Information?"). For convenience these scores were then transformed so that they had a range from 0 - 100.[1]

In the upper panel of Figure 3 we have plotted, subject-by-subject, the mean composite dangerousness scores from the two A + V coders. The actual scores given by the two coders are also shown by small bars joined together. From these individual scores we see that agreement was very high for some subjects (eg., 5, 6, 3) but not so high for others (eg., 1, 2, 8). However, overall, the Pearson Product Moment correlation was acceptable (+0.76) and reliable ($p < .01$)[2].

1. The reader will recognize that this is rather approximate way of completing this initial analysis of the data. Dr. R. Langevin advises us that we should use weightings from the factor analysis (or the large N = 200 study) in order to achieve a more representative overall score.

2. All tests are one-tailed.

In the second panel corresponding data are presented for the Ao coder pair. We see by inspection that the overall correlation between the two coders, though not as high as between the A + V pair ($+ 0.59$, $p < .05$), is reasonably acceptable. We note moreover that there is an appreciable degree of correspondence between the mean of the A + V coder pair and the mean of the Ao coder pair ($r = +0.69$, $p < .02$). Data for the Vo pair are generally similar to the Ao pair with the correlation between the two coders being $+0.62$ ($p < .05$). There was also quite good correspondence between the mean of the Vo pair and the mean of the A + V pair ($r = +0.74$, $p < .02$).

We can now compare data from the psychiatrist-interviewer with those from the three coder pairs. First we should note that the psychiatrist, though reasonably in step overall with the non-clinical raters, offered appreciably lower scores (i.e., he was less inclined to impute dangerousness to the patients than were the clinicians). Second, it is apparent that he was more in accord with the Ao coder pair ($r = 0.72$, $p < .02$) than the other two pairs (A + V - $r = +0.47$, $p = < .01$; Vo - $r = +0.40$, $p < .15$). This is due partly to the fact that he, like the Ao coders, saw Subject 7 as being the most dangerous individual among the nine persons assessed.

We can conclude by stating that in this small study we were apparently able to confirm our earlier finding of acceptable reliability. That we were able to do this both with different coders[1] and a small

1. One of the coders served as an external rater in the main live-interview project.

number of subjects gives us encouragement to refine and develop further the DBRS. It does seem that we can isolate features of "dangerousness" and that we can agree as to whether or not those features are present or absent. Of course, we do not at present know whether or not our scale has validity. This can only be determined when we have follow-up data in hand.

On the basis of the limited data presented here we stress the overall similarity between the responses of the coders and the psychiatrist interviewer. That we should have achieved such correspondence is perhaps surprising in view of the fact that rating conditions were somewhat different (i.e., the psychiatrist rated once at the end of his live interview whereas the coders rated from tape four times).

Our aim in this study, as mentioned earlier, was modest. We set out to test again the reliability of our DBRS. Some success was achieved. It now remains for us to refine the statistical analyses of the data we have in hand and to answer other questions permitted by the design of the study.

REFERENCES

- Sepejak, D. Predicting Dangerous Behaviour: A Review of the Literature. METFORS Working Paper in Forensic Psychiatry, #17.
- Slomen, D., Webster, C.D., Butler, B.T., Jensen, F.A.S., Turrall, G.M., Pepper, J., Penfold, M., Sepejak, D.S., Loftus, L., Byers, D., Chapeskie, T., Mahabir, R.J., Schlager, M., Beckett, K., Shinkoda, A., McDonald, A., Glasberg, R., Jackson, M., Allgood, R., Harman, R., Keeling, K., Taylor, C., Murray, M., Farquharson, D., Lawson, I., Hermanstye, L., and Bendall, L. Assessment of Dangerous Behaviour: Two New Scales. METFORS Working Paper in Forensic Psychiatry, #14, 1979.
- Webster, C.D., Butler, B.T., Jensen, F.A.S., and Turrall, G.M. Constructing Interview-Based Models for the Assessment and Prediction of Dangerous Behaviour: 1, Notes on the Dimensions of the Problem and Some Suggested Criteria against which New Systems might be Evaluated. METFORS Working Paper in Forensic Psychiatry, No.3, 1978 (16 pgs).
- Webster, C.D., Slomen, D., Butler, B.T., Jensen, F.A.S., and Turrall G.M. Constructing Interview-Based Models for the Assessment and Prediction of Dangerous Behaviour: 2, Notes on Recording Interviews and a Discussion of Methodological Possibilities. METFORS Working Paper in Forensic Psychiatry, No.4, 1978 (41 pgs).

FIGURE 1

EXPERIMENTAL DESIGN

CONDITION	CODER	BLOCKS			
		1	2	3	4
A + V	C1 C2	A + V	A + V	A + V	A + V
Ao	C3 C4	Ao	Ao	Ao	A + V
Vo	C5 C6	Vo	Vo	Vo	A + V

A + V = Audio + Video

Ao = Audio Only

Vo = Video Only

PERCENTAGE CODERS' DEFINITE RESPONSES

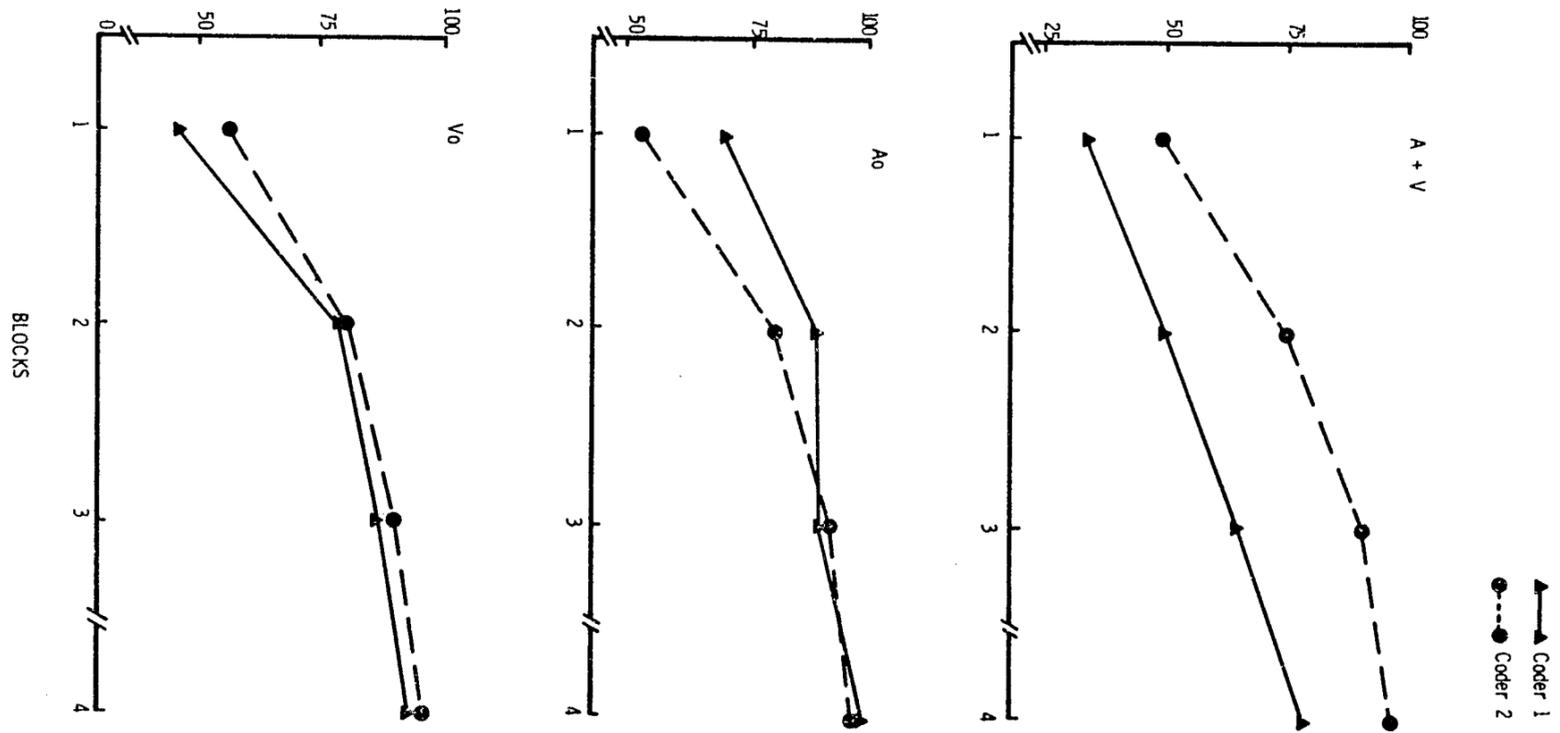


FIGURE 2

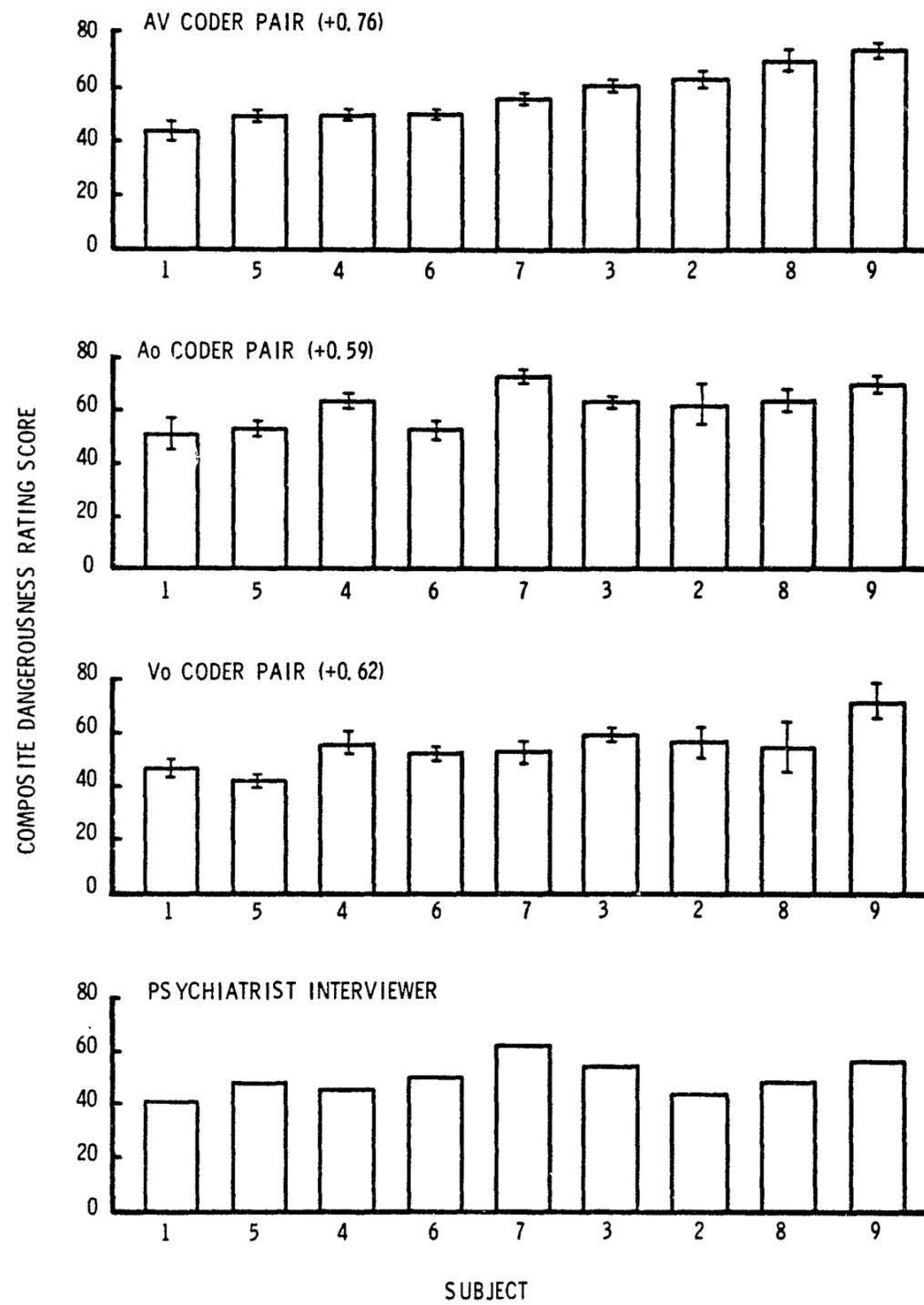


FIGURE 3

END