

- stereotypy. *American Journal of Psychology*, 1958, 71, 50-73.
- LOVAAS, O. I. The relationship of induced muscular tension, tension level, and manifest anxiety in learning. *Journal of Experimental Psychology*, 1960, 59, 145-152.
- MALMO, R. Activation: A neuropsychological dimension. *Psychological Review*, 1959, 66, 367-386.

- STAUFFACHER, J. C. The effect of induced muscular tension upon various phases of the learning process. *Journal of Experimental Psychology*, 1937, 21, 26-46.
- STROUD, J. B. The role of muscular tensions in stylus maze learning. *Journal of Experimental Psychology*, 1931, 14, 606-631.

(Received August 16, 1963)

Journal of Personality and Social Psychology
1965, Vol. 1, No. 5, 510-517

PREDICTING THE SUCCESS OF PEACE CORPS VOLUNTEERS IN NIGERIA¹

WALTER MISCHEL
Stanford University

The outcomes of assessment in 1 of the 1st Peace Corps projects are described. Self-report measures of various ratings during training in the U.S. were obtained on a group of 41 Peace Corps volunteers (PCVs) assigned to teaching posts in Nigeria. The criterion consisted of staff ratings of the volunteers in Nigeria. Scores on self-report measures (F, Ego Strength, and Manifest Anxiety scales) were related significantly to performance abroad. In contrast, global ratings (by training faculty, assessment board, and interviewer), as well as grades and peer ratings, were not significantly related to the criterion. Outcomes were examined also for PCVs with extreme scores on single and combined self-report measures. Results indicate that the self-report measures employed may have utility for screening and selection decisions. Implications for future assessment are discussed.

This paper reports on certain aspects of assessment in one of the first Peace Corps projects. The measures selected were largely determined by practical considerations. There was little time for assessment, and testing had to be inserted into already overcrowded training schedules. Although the lessons of past large scale assessment projects (for example, Stern, Stein, & Bloom, 1956) were available, for the most part they could not be applied. No volunteers were abroad yet, and the details of the volunteers' assignments and probable situations abroad were not known clearly. In brief, the definitions of the criterion performance remained ambiguous and global. Further, because of training

¹ A modified version of this paper was read at the conference "Peace Corps and the Behavioral Sciences," Department of State, Washington, D. C., March 5, 1963. Grateful acknowledgment is due to Alex Sergienko who was responsible for collecting data (with the exception of the self-report questionnaires and objective tests) during training and to David Seeley who collected the criterion ratings in Nigeria. This paper is the responsibility of the author and should not be construed as an official Peace Corps document.

policies and overcrowded time schedules, it was impossible to observe and rate volunteers in training situations sampling the kinds of behaviors required abroad, for example, teaching. These factors imposed severe limitations upon assessment possibilities, but since the assessment situation was so novel and important the opportunity to attempt a prediction study, albeit based on limited data, had to be accepted.

This study describes the use of some structured self-report measures for predicting the performance of volunteers in the Harvard-Nigeria Peace Corps project. The paper also presents results from the use of grades during training, peer ratings, interviews, faculty ratings, and final assessment-board ratings, all collected independently by the training staff during the Harvard training period.

Findings from previous assessment projects led to the following expectations. First, it was anticipated that self-report measures would be of value for broad screening-out purposes, with the poorest scores predicting relatively unsatisfactory performance. Second, it was expected that academic records (grades) obtained during training, and peer ratings would be moderately re-

lated to criterion performance. In view of their past performance, the writer had little optimism with respect to the "global" judgment techniques, namely, the ratings from unstructured interviews, faculty judges, and the assessment conference. However, due to the "face validity" of these latter measures they were included by the training staff and provided an opportunity for comparing the predictive values of a variety of techniques.

METHOD

Subjects

When tested this group consisted of 28 male and 12 female Peace Corps Volunteers (PCVs) trained for approximately 2 months at Harvard University during the summer of 1961 in preparation for teaching assignments in the schools of western Nigeria. The candidates in training were preselected from a much larger national pool of applicants. Prior to inclusion in the training program, candidates were screened in Washington, D. C., by an assessment staff which based the screening decisions primarily on the PCV application questionnaire, letters of recommendation, the applicant's previous academic and work record, and performance on the Peace Corps entrance examinations.

From this group of 41, 4 were screened out at the end of the Harvard training program. After a further training period at the University of Ibadan (Nigeria) the PCVs took their teaching assignments in different schools throughout the western region of Nigeria. At the time of the criterion ratings (April 1962), 35 PCVs remained in Nigeria.

Predictor Data: Self-Reports

The self-report data were collected by the writer in late August of 1961, that is, towards the end of the Harvard training program. The tests were administered to all PCVs in a group session at Harvard. It was made clear to the PCVs that these tests were used for research purposes only and it was stressed that the results would not affect any Peace Corps decisions that might be made about individual respondents. (This was indeed the case.) The tests were administered in the standard ways customary for each instrument. The test battery consisted of several paper-and-pencil measures. These included the 24-item version of the California F Scale composed of 12 items from the original F Scale (Adorno, Frenkel-Brunswick, Levinson, & Sanford, 1950) and 12 items dealing with "traditional family ideology" (Levinson & Huffman, 1955), the Ego Strength (*Es*) scale (Barron, 1953), the Manifest Anxiety (*MA*) scale (Taylor, 1953), and an abbreviated intelligence test.² Items from the *Es* and *MA* were interspersed within the same format.

A group-administered adaptation of three verbal tests (Similarities, Information, Vocabulary) from the Wechsler Adult Intelligence Scale (WAIS) was administered. The correlation between total

The above measures were selected because they can be administered and scored quickly, simply, and objectively, and a considerable research literature exists for each of them. Most important, it was known that the criterion ratings would be fairly global and the assessors had insufficient information concerning the details of the criterion to make specific predictions. Therefore, the predictions of relative success (or relative failure) had to be based on fairly general considerations, in accord with the broad goals stated by the Peace Corps. These goals included the PCVs' appreciation of the new culture and the development of mutual understanding, as well as competent job performance. Finally, the same scales were included in several Peace Corps projects and thus make cross-project comparisons ultimately possible.

Predictor Data: Measures from the Training Staff

The second set of available data came from the training staff. These data were collected independently by the training staff at Harvard during the summer of 1961 and included the PCVs' academic grades during training, faculty evaluations of the PCVs, peer ratings, interview ratings, and assignment to a final category by the assessment review board at the end of training.

The index of academic performance during training was the PCV's grade record. Specifically, the datum for each PCV was his mean grade during training.

The peer ratings, obtained near the end of the training period, required each PCV to list the seven persons from the group with whom he ideally would most like to share his Nigerian school assignment. More specifically, each PCV indicated the one person (A) with whom he would ideally like to be assigned as a school colleague in Nigeria; then the PCV was asked to list two additional others (B); followed by four additional others (C). Crude "weighted peer ratings" were computed in which each A nomination was assigned a value of 3, each B nomination a value of 2, and each C nomination a value of 1. For each PCV the sum of these scores was computed.

There were 10 faculty raters, although not each faculty person rated each PCV. Only faculty members on the training staff who felt that they knew the PCV sufficiently made these ratings and an average of six faculty ratings was obtained for each PCV. Each rater assigned a score on a global rating scale from 1 to 5 (ranging from "outstanding" to "must not go") and the data are based on the mean faculty rating for each PCV.

Each PCV was individually interviewed for about 30 minutes. The interview was essentially unstructured and "informal," "open-ended," and ranged

scores on these three subtests and total criterion points was so low ($r = .002$) that no subtler analyses were attempted. This finding must be interpreted cautiously since the administration of the WAIS was drastically deviant from the standard procedure.

over many topics including "motivation" for joining the Peace Corps. It was conducted by an educator with counseling experience who served in a counseling and liaison role and who remained essentially "in residence" with the PCVs during the training program at Harvard. The interviewer recorded his impressions on a 5-point rating scale, reflecting his judgment of the PCV's probability for successful teaching performance in Nigeria. These ratings appear to have been based on a global impression and were probably influenced by observations outside the interview as well as by the content of the interview itself.

At the end of training each PCV was recommended for one of five categories by an assessment board subcommittee, consisting of three staff persons, all of whom had occasion to observe the PCVs during training. These recommendations were made by the group on the basis of their discussion and review of the data obtained during training (faculty evaluations, academic records, peer ratings, and interview). The categories were: I, outstanding ($N = 4$); II, very good ($N = 17$); III, satisfactory ($N = 11$); IV, some doubt ($N = 7$); V, not recommended ($N = 2$). These recommendations were discussed in detail by a larger final assessment review board, representing a variety of disciplines and coming with diverse assessment experiences. The final review board discussed each PCV individually and had available a descriptive summary prepared by the subcommittee for each PCV. The subcommittee recommendations were accepted, in essence, except that two PCVs from Category IV were moved into Category V. Thus, of the 41 candidates, 4 were asked not to continue at the end of training. Two kinds of data analyses were computed for these category assignments: correlations between the category assignment and the criterion scores, and chi-square tests comparing combined Category I and II PCVs with Category III and IV PCVs on performance above or below the criterion median.

The training staff had no knowledge of the PCV's scores on the objectively scored tests administered by the writer. Thus, results from the measures described above under Predictor Data: Self-Reports did not affect selection decisions, whereas the data obtained from the techniques described under Predictor Data: Measures from the Training Staff did affect such decisions at the end of training.

There may have been important differences in the PCVs' test-taking "set" for the measures described above in the self-report data opposed to those described in the training staff data. The measures under the former were administered by an examiner unfamiliar to the PCVs and with the implication that the results would not affect decisions about the candidates. In contrast, the measures under the latter were administered by persons associated with the training program and with the implication that the outcomes could affect decisions about the candidates. Hence, it is to be expected that responses to the latter measures might be much more subject to defensive distortions, attempts to "fake good," etc.

Predictions

It was predicted that strong "authoritarianism" as measured by high scores on the F Scale, and low "ego strength" as measured by low scores on the Es scale, would be indices of poor performance in Nigeria. That is, it was expected that the strong presence of "authoritarianism" associated with high F Scale scores and the "low ego strength" associated with poor Es scores would impede the PCV's effectiveness in a very different, foreign environment, under strenuous personal and interpersonal circumstances. With respect to the MA scale, it was anticipated that extremely high scores would be an index of debilitating "anxiety" and that PCVs with such extreme scores would not function as effectively as PCVs with more moderate scores.

Especially since this was a prescreened group, strong correlations between the predictors and criterion performance were not expected. Rather, it was anticipated, and explicitly predicted, that the greatest value of these self-report scales would be for the prediction of relatively poor performance on the part of PCVs with relatively extreme "poor" scores on the self-report measures. The a priori cutoff point arbitrarily selected for predicting poor performance was the "poorest" 25% of the distribution of scores on each scale. Thus it was predicted that PCVs with scores in the highest 25% on the F or MA scales, or in the lowest 25% of the distribution on the Es scale, would be the "doubtful" candidates. These predictions were formally recorded in September 1961, and were not made available to the training or field staff.

As regards the measures collected by the training staff, it was anticipated that performance during the training periods, as measured by grades and peer ratings, would relate positively to performance in Nigeria. In view of the fact that the unstructured or "global" interview has had a persistent history of unsatisfactory predictive validity in previous assessment projects, no positive relationship was anticipated. Similarly, the global category assignments by the assessment board and the global ratings of faculty judges were not expected to yield useful predictions.

Criterion Ratings

Ratings of the relative success of each PCV were obtained in April 1962. These are the criterion ratings which were made independently by the field staff in Nigeria and without any knowledge of the predictor data collected earlier. The criterion ratings were made on six dimensions, each of which was described in considerable detail by the field staff. The dimensions were: teaching effectiveness; person-to-person contacts (interpersonal behavior with Nigerians); appreciation of Nigerian culture; representation of American culture by (a) personal behavior and (b) by interpreting American culture to Nigerians; and adjustment to the assignment. Teaching effectiveness was given extra weight and rated on a 50-point scale, whereas the other five ratings were made on 10-point scales. The total criterion points (sum of

all ratings) ranged from 54 to 87, with numerous tied scores. The ratings were submitted by the field staff supervisor and represent his collation and pooling of judgments as described below. The ratings of different judges were not individually reported and therefore the reliability of the criterion ratings cannot be evaluated. In effect, the criterion data must be treated as the ratings of a single judge, that is, the field supervisor.

The following quotation, excerpted from the May 31, 1962, interim report by the field supervisor for the Harvard-Ibadan Peace Corps Training project highlights the serious limitations of the ratings. According to this report, the volunteers were rated:

on the various elements of the desired PCV behavior It does not purport to be accurate. The ratings are in many instances based on very scant evidence and in some cases amount to nothing more than guesses. Much of what the volunteers have been doing since arriving at their posts has gone unobserved by any Peace Corps or training staff, and the judgments recorded here have been constructed out of the fragments of evidence we do have and the second-hand opinions of school principals and others who have had more opportunity to observe the volunteers on the job. [Nevertheless:] If . . . [the poorest scores] could have been predicted ahead of time, it would have been better not to send these volunteers.

RESULTS

Correlations between the six criterion subscales and the total criterion scores ranged from .72 (person-to-person contacts) to .85 (teaching effectiveness and adjustment to assignment). In view of these reasonably high correlations, as well as the global nature of the criterion ratings themselves, all data analyses are based on the total criterion scores.

Table 1 presents the intercorrelations between the predictors for the total sample of PCVs for whom predictor data were available ($N = 41$). Note that the correlations between each self-

report measure and all other measures reach statistical significance in only one instance, namely, the negative relationship ($r = -.58$) between Es and MA. Scores on the self-report measures are not related to scores on the measures collected by the training staff. In contrast, note with respect to the latter data that all the intercorrelations with only two exceptions are statistically significant, although there is a wide range of magnitude. The only nonsignificant correlations are between academic record during training and the interview and between the former and peer ratings.

The highest correlations are between assessment-board categories and faculty ratings ($r = .82$) and between assessment-board categories and the interview ($r = .72$). Since the assessment-board category ratings were made at the end of training, with all data from the training-staff predictors available, it is of interest to see that when the assessors made their final ratings they were apparently most influenced by the results from the faculty evaluations and the interview. This is consistent with the frequently noted tendencies of judges to rely most heavily on data from other judges who are most closely their peers (rather than the ratings of the PCV peers or on the academic grades of the PCVs during training). The negative consequences of this will become apparent when the correlations between the predictors and the criterion are examined. In other words, the final global decisions made about PCVs at the end of training were most strongly related to faculty ratings and the interview (although the correlations with peer ratings and academic grades were also significant). The correlations between the self-reports and data collected by the training staff in no case approach statistical significance. In sum, there is no significant relationship between scores on the self-reports and the measures ob-

TABLE 1
INTERCORRELATIONS BETWEEN PREDICTORS ($N = 41$)

	Self-reports			Academic grades	Peer ratings	Faculty ratings	Interview
	F	Es	MA				
MA	-.12						
Academic grades	.03	-.58**					
Peer ratings	-.13	.09	-.03				
Faculty ratings	-.20	.12	-.13	.20			
Interview	.01	.15	.10	.55**	-.35*		
Assessment board	-.13	.14	.18	.25	.35*	.61**	
	-.15	.04	.22	.55**	.34*	.84**	.72**

* $p < .05$.
** $p < .01$.

TABLE 2
CORRELATIONS BETWEEN PREDICTORS AND TOTAL CRITERION POINTS (N = 35)

Criterion	Predictors							
	Self-reports			Academic grades	Peer ratings	Faculty ratings	Interview	Assessment board
	F	Es	MA					
	-.45**	.34*	-.34*	.24	.29	.27	.13	.20

* $p < .05$.
** $p < .01$.

tained by the training staff, although there are reasonably high correlations between most of the latter measures.

Table 2 shows the correlations between predictors and criterion ratings. The correlations are based on the 35 PCVs in Nigeria at the time of the criterion ratings.³ Inspection of Table 2 reveals that scores on each of the three self-report measures are significantly related in the expected direction to criterion ratings.⁴ In contrast, for the data gathered by the training staff, none of the correlations reach significance. The lowest correlation is for the interview; the highest correlations are for peer ratings ($r = .29$), faculty ratings ($r = .27$), and academic grades ($r = .24$), but all of these fall short of acceptable significance levels, with $r = .33$ required for $p < .05$.⁵

³ All tests of the relationships between predictors and the criterion were based only on the 35 PCVs remaining in the field at the time of the criterion rating. Since decisions about the 4 PCVs screened out at the end of training were based on some of the predictor measures these cases cannot be included in the analyses dealing with the interrelationships between predictors and outcomes. Of the 2 PCVs who were "dropouts" in Nigeria, 1 scored in the poorest quartile of the F Scale and the other in the poorest quartile of the Es scale.

⁴ Inspection of the relationships between predictors and criterion ratings for males and females separately revealed no strong sex differences and the data for males and females were combined in all statistical tests.

⁵ Since the assessment-board ratings resulted in categories varying markedly in size of N , a contingency table was tabulated in addition to the Pearson correlation between categories and criterion ratings. This consisted of comparing the outcomes (above or below the criterion median) for PCVs with "high" as opposed to "low" assessment-board ratings, with these ratings dichotomized as closely as possible into two groups of approximately equal size. Of the PCVs with "high" assessment-board ratings, 11 were above the criterion and 9 were below it; of the PCVs with "low" ratings, 7 were above and 7 were below the criterion median.

Examination of the "hits" and "misses" obtained if cutoff points on the predictor variables are used to make screening decisions provides an alternative analysis which is more interesting than the correlation patterns. Recall that with respect to the self-report measures it was predicted that a cutoff point, delineating the poorest 25% of the distribution, would provide an effective index of poor or doubtful criterion performance. To differentiate PCVs in terms of good or adequate as opposed to poor or doubtful performance the distribution of criterion points was dichotomized at the median. This crude dichotomy of criterion performance seems justified in view of the grossness of the criterion measure and provides at least a rough estimate of relatively good as opposed to relatively poorer performance in the field.

Table 3 presents the outcomes if cutoff points are used to delineate the poorest 25%, 33%, and 50% of the score distribution on each self-report (that is, high scores on F and MA, low scores on Es). The table shows the number of PCVs who would have been "erroneously screened out" (that is, performance above the criterion median) and the number "accurately

TABLE 3
N ERRONEOUSLY AND ACCURATELY SCREENED OUT IF CUTOFF POINTS ARE USED TO SCREEN OUT PCVS WITH "POOR" SCORES ON EACH SELF-REPORT MEASURE

Outcome	Cutoff used ^a	Self-reports		
		F	Es	MA
N erroneously screened out (above criterion median, N = 18)	25%	1	1	1
	33%	1	3	4
	50%	2	7	6
N accurately screened out (below criterion median, N = 17)	25%	8	7	7
	33%	10	8	8
	50%	14	11	10

^a "Poorest" 25%, 33%, and 50% of scores on each self-report measure.

TABLE 4
CRITERION PERFORMANCE OF PCVS WITH "POOR" OR "GOOD" SCORES ON TWO AND ON ALL THREE SELF-REPORT MEASURES

Criterion performance	Self-report scores							
	F and Es		F and MA		Es and MA		F, Es, and MA	
	Poor	Good	Poor	Good	Poor	Good	Poor	Good
Above median (N = 18)	0	8	0	7	4	7	0	6
Below median (N = 17)	9	1	7	1	8	4	6	0

Note.—"Poor" scores are scores above the median on F or MA and below the median on Es; "good" scores are scores below the median on F or MA and above the median on Es.

screened out" (performance below the criterion median) for the closest approximation of each of these cutoffs.

Table 3 clarifies two points. First, note that for the 25% cutoff, the three self-report measures, and especially the F Scale, seem to make useful discriminations. With this cutoff the F Scale would have screened out accurately eight relatively poor PCVs. Only one PCV above the criterion median would have been screened out erroneously. With the Es and MA scale similar results would have been obtained, although slightly less clearly.

Second, obviously as the cutoff point is changed to include an increasingly large percentage of the sample, the number of PCVs erroneously screened out increases greatly. The notable exception is the F Scale, on which a cutoff of even 50% would have resulted in only 2 PCVs being erroneously screened out, whereas 14 of the 17 persons below the criterion median would have been identified accurately.

All the other predictor variables were examined in a similar way. The number of PCVs who would have been accurately screened out never exceeded twice the number erroneously screened out and in most instances the number erroneously screened out equaled or approached the number accurately screened out.

Outcomes were also examined for PCVs who had poor scores on two or on all three self-report measures. The distribution on each self-report measure was dichotomized into scores above and below the median. "Poor" scores were defined as scores above the median on F or MA and below the median on Es. Table 4 presents outcomes for PCVs with poor scores on any two self-report measures and on all three self-reports. The main advantage of using such combinations of poor scores is that the number of PCVs erroneously screened out is reduced to zero for

all combinations in which the F Scale is included.

Post hoc, it is of interest that delineating PCVs who attained "good" scores on two or on all three self-report measures would also have had some usefulness for selecting PCVs into the project. Good scores were defined as scores below the median on F or MA and above the median on Es. Table 4 includes the outcomes for PCVs with good scores on two or on all three self-reports. Self-report combinations which include the F Scale would have been useful but the combination of Es and MA would not have been valuable. Further, it would not have been very fruitful to delineate PCVs with good scores on each self-report measure separately. For example, on the F Scale, the quartile of PCVs with the best (that is, lowest) scores contained eight volunteers and of these, six were above the criterion median but two were below the median and this was the best yield obtained from the use of extreme good scores on any single self-report.

DISCUSSION

The overall results are consistent with the findings of other assessment projects in which both the assesses and the criterion were drastically different (for example, posthospital adjustment of schizophrenics). For example, many of the conclusions reached by Marks, Stauffacher, and Lyle (1963), in a recent effort to predict outcomes in schizophrenia, fit the present study. These authors conclude that "the best predictors of year-end adjustment are those which can be most expeditiously obtained [p. 126]." In their study, as in the present report, one of the best predictors was the "old 'impure' F" Scale.

The failure of "global" techniques, specifically the interview and global ratings by staff judges and the assessment board, is noteworthy and stands in contrast to the relative utility of simple self-report techniques. Again, this is consistent

with past assessment experiences (Cronbach, 1956). In general, the results support the usefulness of simple self-report inventories and point, once more, to the dangers of relying on global ratings (the unstructured interviews and staff judges) and on nonquantitative interpretations of multiple data (the assessment-board ratings). These latter techniques are usually the most costly, time-consuming efforts in the assessment program. Although they seem compelling and valid to those who use them, the evidence supporting their utility remains minimal. As Table 1 showed, the assessment-board ratings, which determined the decisions about volunteers at the end of training, were most strongly related to the faculty ratings and the interview—and the interview turned out to be the measure with the lowest predictive validity. Staff judges (faculty ratings, interview, assessment boards) showed considerable agreement with each other—but their judgments were not significantly related to criterion performance.

The correlations between criterion ratings and academic grades, and ratings by peers, were also not significant, although the magnitude of correlations tended to be somewhat higher than those obtained from the interview and the assessment-board categories. It should be recognized that the peer-rating technique employed was extremely gross, and that at the time of the peer ratings the PCVs had very little knowledge about the field conditions which awaited them abroad. With respect to grades during training it is quite possible that the behaviors which determined these grades had little relevance for actual performance abroad. Indeed, informal comments from the field indicated that in many instances the content of the academic training program and the actual demands and conditions of the field experience were at best remotely related.

Caution must also be applied when interpreting the results from the self-report measures. In spite of "significant correlations," their utility for assessment decisions remains dubious. However, these correlations are more impressive in view of the fact that the sample of PCVs was prescreened and consequently there was a relatively narrow range of scores on each self-report measure. For example, on the F Scale, on which the highest score could be 168, actual scores ranged from 25 to 82.

"Screening-out" decisions, based on extremely poor self-report scores (for example, the poorest quartile), especially with the F Scale, would have reduced appreciably the number of poor candidates sent abroad. But obviously the total "costs"

of such decisions are complicated by many considerations.

Replication of the present findings in other Peace Corps projects would have increased faith in the results of the present study. However, findings for another project in Ghana already show a failure to replicate (Smith & Roth).⁶ Indeed, although many of the predictor measures used in Ghana were similar to those in the present study, none were significantly related to criterion ratings. Although it is disappointing that the positive results obtained from self-reports were not replicated, this may have been due to diverse reasons, including differences between the present sample and the Ghana sample with respect to the PCVs selected, the job requirements and field conditions, the determinants of criterion ratings and the PCVs' test-taking "set."

A possible difference in the test-taking "set" of PCVs on the self-reports in the two samples may be of special importance. Recall that in the present project the self-reports were administered explicitly as part of a research project with the promise that results would not determine decisions about individual PCVs. The set may have minimized distortions in test responses and enhanced the obtained relationships with criterion ratings. Current research by the writer on "self-prediction" under conditions in which accurate self-reports do not lead to negative consequences for the respondent suggest that self-prediction can provide remarkably accurate outcome indices (Mischel, 1963). For example, when the data were collected under a "for research only" set, correlations between self-predicted course grades and obtained grades exceeded .60 for a sample of 90 undergraduates. Under appropriate test-taking sets self-report measures and self-predictions may be valuable techniques for helping candidates in their "own" "self-selection" process. Obviously this could only be effective if the relevant criteria are described to the respondent and if he is presented with sufficient alternatives so that an accurate response does not result in detrimental consequences for him. For example, in connection with the Peace Corps self-reports as a basis for self-selection might be an effective technique if alternative assignments, requiring different attributes, are offered and described in terms which have equivalent social desirability.

Although many psychologists are privately convinced that "if you want to know how a person is going to behave, ask him," research into the

⁶ "Ghana Project," progress report to the Institute of Human Development, University of California, Berkeley, April 1963. (Ditto)

conditions under which accurate self-prediction is and is not possible remains minimal. Particularly in view of the persistent difficulties encountered when "expert judges" attempt to predict the behavior of others, a renewed exploration of the self-report seems worthwhile, not merely to obtain the subject's self-descriptions but also to obtain his explicit outcome predictions for specific contingencies.

REFERENCES

- BOGNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
- BARRON, F. An ego strength scale which predicts response to psychotherapy. *Journal of Consulting Psychology*, 1953, 17, 327-333.

Journal of Personality and Social Psychology
1955, Vol. 1, No. 5, 517-520

SOME VERBAL ASPECTS OF PRIMARY-PROCESS THOUGHT:

A PARTIAL REPLICATION¹

STANLEY BLUMBERG

AND

DONALD W. GILLER²

University of Texas Southwestern Medical School

Terrell State Hospital, Texas

This study compared the conceptual performance of chronic schizophrenics with acute and nonschizophrenic psychiatric patients on a word task requiring the selection of synonyms. All groups showed evidence of distractibility or primary-process thinking by choosing antonyms and homonyms on some of the items. It was concluded that such cognitive errors are not unique to schizophrenia, but occur in all groups to an extent dependent upon severity of illness.

The present study is an approximate replication of an experiment demonstrating some verbal aspects of primary-process thinking in schizophrenia (Burstein, 1961). Among the attributes of primary-process thought Freud (1956) described the confusion of opposites with identities and similar word sounds with similar meanings. Based on these descriptions, Burstein constructed two multiple-choice word tasks to put Freud's formulation to an empirical test. Using a group of long-term hospitalized schizophrenics and a group of normals, Burstein found that, as predicted, the schizophrenics more often did confuse opposites with identities, selecting antonyms as

¹We are grateful to A. G. Burstein for his generosity in permitting us to use his antonym and synonym word lists.

²Now at the Dallas State Adult Mental Health Clinic.

- CRONBACH, L. J. Assessment of individual difference. *Annual Review of Psychology*, 1956, 7, 173-196.
- LEVINSON, D. J., & HUFFMAN, P. E. Traditional family ideology and its relation to personality. *Journal of Personality*, 1955, 23, 251-273.
- MARKS, J., STAUFFACHER, J. C., & LYLE, C. Predicting outcome in schizophrenia. *Journal of Abnormal and Social Psychology*, 1963, 66, 117-127.
- MISCHEL, W. *Studies in self-prediction*. Stanford: Stanford University, 1963. (Mimeo)
- STERN, G. G., STEIN, M. I., & BLOOM, B. S. *Methods in personality assessment*. Glencoe, Ill.: Free Press, 1956.
- TAYLOR, JANET A. A personality scale of manifest anxiety. *Journal of Abnormal and Social Psychology*, 1953, 48, 285-290.

(Received August 28, 1963)

synonyms, and similar word sounds with similar meanings, choosing homonyms for synonyms.

Burstein also found a significant difference in mean IQ scores between the two groups, in spite of their having been matched for age, education, and estimated "premorbid intelligence." He interpreted these results in terms of the schizophrenic pathology, that is, with increasing pathology there is increased primary-process thinking and intellectual deficit. Evidence for this hypothesis was the significant correlation between scores on the word tasks and the intelligence measures for the schizophrenic group, with no significant correlation for the normal group.

One question unanswered in Burstein's study is whether this specific error tendency of confusing antonyms and homonyms with synonyms is uniquely schizophrenic or is common to psychopathology in general. Chapman (1958) found