

Mental Health Screening: Utility of the Psychological Screening Inventory

Richard I. Lanyon
Arizona State University

The Psychological Screening Inventory (PSI) was designed for use as a rapid mental health screening device to identify persons who should receive more extensive evaluation. This study evaluated the accuracy of the three screening scales using receiver operating characteristic (ROC), positive predictive power–negative predictive power, and sensitivity–specificity analyses. Persons with major psychiatric disorder ($N = 536$) were distinguished from control participants by the Alienation scale with an area underneath the ROC curve (AUC) of .905; persons with significant antisocial behavior ($N = 410$), by the Social Nonconformity (Sn) scale with an AUC of .811; and persons with general psychological distress ($N = 620$), by the Discomfort (Di) scale with an AUC of .773. There were comparable findings for the other comparison procedures, and some results were better for men and women separately. An empirically developed index (the Overall Screening Index—Revised) showed significant improvement over the Di scale in identifying persons with general psychological distress (AUC = .842). These findings support the utility of the PSI for its intended purpose.

Keywords: screening, mental health, Psychological Screening Inventory, PSI

The Psychological Screening Inventory (PSI; Lanyon, 1970, 1973, 1978, 1993, 2005), first published in 1973, was developed to meet the need for a brief mental health assessment device when time and personnel are at a premium. It was intended primarily for screening purposes, to be used in identifying persons for whom a more complete evaluation might be indicated. Designed to be completed in about 15 minutes, the PSI contains 130 true–false items over five scales, including three basic screening scales.

The names of the three screening scales were deliberately chosen to downplay the significance of the scales to the casual observer. The Alienation (Al) scale was designed to screen for major psychiatric disorder, the Social Nonconformity (Sn) scale for significant antisocial behavior, and the Discomfort (Di) scale for general psychological distress. These concepts were originally chosen to be loosely allied with the three traditional psychiatric categories of psychosis, character disorder, and neurosis. The remaining two scales were intended to enhance

the general usefulness of the instrument. The Expression (Ex) scale represents the basic personality dimension of extraversion–introversion or undercontrol–overcontrol, and the Defensiveness (De) scale was designed to assess degree of personal defensiveness in responding to the items. Some initial empirical data supporting the use of the Al scale in discriminating psychiatric inpatients from control participants, and the use of the Sn scale in discriminating incarcerated persons from control participants, were given in the *PSI Manual* (Lanyon, 1973, 1978).

The purpose of the present study was to conduct a comprehensive evaluation of the empirical accuracy of all three PSI screening scales in making the discriminations they were designed to make. As a basis for comparison, briefer evaluations were also made of the three scales in making the remaining two discriminations.

The study also evaluated the accuracy of an empirically developed index for identifying general mental health problem status. This index was developed by Overall (1974), who contrasted the PSI scores of a psychiatric outpatient group with those of the control participants mentioned above (the original PSI normative group, $N = 1,000$; see Lanyon, 1970, 1978). Studying men and women together, Overall

Correspondence concerning this article should be addressed to Richard I. Lanyon, Department of Psychology, Arizona State University, Tempe, AZ 85287-1104. E-mail: RLanyon@asu.edu

found that the two groups could be discriminated with greater than 80% accuracy using a discriminant function based on raw scores, giving primary weight to the AI and Di scales plus a contribution from the De scale as a correction for underreporting. The specific function was as follows: $.417AI + .034Sn + .244Di + .046EX + .307De$. (A second discriminant function involving a set of specially constructed factor-based scales performed slightly less well.)

Granted that the accuracy of the discrimination would decrease on cross-validation, Overall's (1974) work is noteworthy because it showed empirical success for the PSI in an outpatient mental health screening context and because the screening algorithm has obvious face validity. In the present study, the index was closely approximated by a much simpler index, also based on raw scores: $5AI + 3Di + 4De$. This approximation, here termed the *Overall Screening Index—Revised* (OSIR), was shown to correlate greater than .99 with the original index for the first 100 of the psychiatric inpatients used in the present study. The approximation was used rather than the original index because it can be quickly computed by hand and therefore can be routinely used in practical settings. Because the OSIR was developed in an outpatient mental health setting, its success was studied here primarily as an alternative to the Di scale.

Each of the three screening scales was predicted to show satisfactory discrimination on the screening task for which it was designed and better discrimination than either of the other two scales. The OSIR was also predicted to serve as an alternative to the Di scale in discriminating persons with general psychological distress from control participants.

Method

Participants

The PSI scores of a total of 2,566 persons were available for this study. Of these, 1,000 were the original participants in the PSI normative group (Lanyon, 1970, 1978). These individuals were representative of the United States in age and education and came from four diverse geographical locations across the country. They are referred to as control participants from

here on, and their scores were used for comparison purposes with three groups of target participants. The participants with major psychiatric disorder ($N = 536$) came from three different psychiatric inpatient settings, the participants with significant antisocial behavior ($N = 410$) came from five different penal institutions and a state hospital, and the participants with general psychological distress ($N = 620$) were outpatients at four different community mental health centers. These sources are identified in Table 1 and corresponding demographic data are given in Table 2.

Procedure

Four approaches to evaluating screening accuracy were used: effect sizes, receiver operating characteristic (ROC) analysis, positive predictive power and negative predictive power (PPP–NPP), and sensitivity at a preselected level of specificity. Because the distributions for men and women differed substantially in some cases, all analyses were done for men and women separately as well as combined.

Effect size is a commonly used measure of the strength of the relationship between two variables. The measure of effect size chosen for the present study was Cohen's d (Cohen, 1988), which is the difference between the means relative to the variation within the groups (Hays, 1994). As a specific example, if the standard deviation for both groups were 10 (an approximation of the present data) and the means differed by 10 points, the effect size would be 1.00. An effect size of .80 or more is considered to be large (Cohen, 1988).

ROC analysis represents an integration of accuracies across the range of possible cut scores or, more specifically, across different levels of sensitivity and specificity (Rice & Harris, 1995; Swets, 1992). The area underneath the curve (AUC) "reflects the probability that a randomly selected person from one population will have a scale score that exceeds that of a randomly selected person from the other population" (Hsu, 2002, p. 414). In more general terms, it can be viewed as the probability that any individual participant will be correctly classified using various cut scores. In a comparison using ROC analysis, the data from each of the two comparison groups are combined to determine the AUC.

Table 1

Sources of the Psychological Screening Inventory Scores Used in the Present Study, Listed According to the Screening Scale of Interest

<i>N</i> (male, female)	Source	Acknowledgment
Al scale (major psychiatric disorder), <i>N</i> = 536		
220 (110, 110)	State hospital, unselected	Jay Lewis
133 (52, 81)	State hospital, unselected	Lauretta Walker
183 (80, 103)	State hospital, schizophrenia	Kalman Csapo
Sn scale (significant antisocial behavior), <i>N</i> = 410		
80 (40, 40)	Young adult reformatory	Richard Lanyon
79 (0, 79)	State prison	Warren Webb
49 (0, 49)	State prison	Anthony Broskowski
20 (20,0)	Young adult correctional	Richard Lanyon
122 (64, 58)	State hospital, antisocial personality disorder	Kalman Csapo
60 (46, 14)	Awaiting criminal trial	Richard Lanyon
Di scale (general psychological distress), <i>N</i> = 620		
620 (310, 310)	Four community mental health centers, unselected	Gary Greer

Note. Al = Alienation; Sn = Social Nonconformity; Di = Discomfort.

Sensitivity refers to the proportion of the target group that is correctly identified by the test or, more precisely, identified by the particular cut score that is used. *Specificity* refers to the proportion of the comparison or control group that is correctly identified. It is recognized that ROC analysis represents an integration of accuracies over the entire range of cut scores. However, because the accuracy data originally reported in the *PSI Manual* (Lanyon, 1973, 1978) for the discriminations to be made by the Al and Sn scales were expressed in terms of percentage accuracy (i.e., sensitivity–specificity), those analyses were also conducted here. Because the original analyses involved specificities of about .8, cut scores representing a specificity level as close as possible to .80 were chosen here for ease of comparison.

Positive predictive power (PPP) is the probability that a person above the cut score selected is actually member of the target group, and *negative predictive power (NPP)* is the probability that a person below the cut score selected is actually a member of the control group. Although PPP and NPP are dependent on particular cut scores and base rates, they have the advantage of relating directly to clinical practice. For example, a PPP of .90 indicates that a person with a score above the cut score is 90% likely to be a member of the target group. In the present analyses, a single base rate of .20 (i.e., 20% target persons) was chosen as being appropriate to many of the situations that would be encountered in the practical use of the PSI. The same cut scores were used as in the sensitivity–specificity analyses.

Table 2

Demographic Data for All Participant Groups

Participant group	Gender		Age		Education ^a	
	Male	Female	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Major psychiatric disorder	242	294	33.7	10.6	10.7	2.5
Significant anti-social behavior	170	240	32.5	9.2	11.2	2.5
General psychological distress	310	310	31.7	10.3		
Normative group	500	500	35.4	13.1	13.5	3.3

^a Education data were not available for the participants with general psychological distress.

Results

Mean *T* scores and standard deviations on the five PSI scales and the OSIR are shown in Table 3 for the four participant groups, for men and women separately and combined. The *T*-score form was used for the individual scales for ease in comparison, and the OSIR data were based on raw scores. Also shown in Table 3 are effect sizes (Cohen's *d*) on the three screening scales and the OSIR for the three target groups when compared with the control group. These data permit direct comparisons among the three scales and the OSIR in making all discriminations.

The four specific sets of analyses are presented next, arranged according to each specific participant group and the scale designed to discriminate that group.

Major Psychiatric Disorder (AI Scale)

Effect size. Effect sizes for the AI scale in distinguishing inpatients from the control group,

shown in Table 3, were very large (1.80 and greater), and much larger than those for the other scales. The single exception was with the OSIR, which showed comparable effect sizes, including the largest effect size in the entire set of analyses (1.96, in distinguishing men with major psychiatric disorder from male controls). The AI scale also showed smaller but still large effect sizes (averaging about 1.1) in distinguishing the other two target groups from control participants.

ROC analysis. Results of the ROC analyses, reported in Table 4, represent the ability of the AI scale to distinguish between the control group and psychiatric inpatients across a broad range of cut scores and base rates. For all participants, the AUC was .905 (95% confidence interval = .889-.921). For men and women separately, the AUCs were .912 and .898, respectively. Hosmer and Lemeshow (2000) suggested a range of .7 through .8 as acceptable discrimination, .8 through .9 as excellent discrimination, and above .9 as outstanding discrimination.

Table 3
Mean *T* Scores and Standard Deviations for the Five Psychological Screening Inventory (PSI) Scales and Raw Scores for the Overall Screening Index-Revised (OSIR) for the Four Participant Groups

Group and gender	Scale															
	AI			Sn			Di			Ex		De		OSIR (raw score)		
	<i>M</i>	<i>SD</i>	<i>d</i>	<i>M</i>	<i>SD</i>	<i>d</i>	<i>M</i>	<i>SD</i>	<i>d</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>d</i>
Major psychiatric disorder																
M	70.2	10.7	1.89	53.4	11.7	0.32	61.0	12.6	1.01	44.9	8.4	51.3	10.7	146.7	30.0	1.96
F	69.0	10.5	1.80	55.0	9.9	0.48	56.5	11.5	0.61	47.9	8.5	48.9	11.8	146.2	28.4	1.52
M&F	69.6	10.8	1.81	54.0	10.6	0.39	58.7	13.3	0.77	46.3	8.3	50.0	12.8	146.4	29.1	1.67
Significant antisocial behavior																
M	62.2	13.9	1.10	65.1	10.8	1.49	57.3	12.8	0.68	47.5	8.5	47.5	11.3	124.3	32.3	1.06
F	60.1	11.3	1.06	64.3	11.4	1.36	56.6	11.3	0.63	50.8	8.9	47.7	12.2	130.6	30.7	0.87
M&F	61.1	12.9	1.01	64.7	12.1	1.43	56.9	12.8	0.63	49.4	8.7	47.6	11.0	128.0	31.5	0.96
General psychological distress																
M	62.5	12.6	1.13	54.0	10.4	0.39	61.0	12.5	1.00	45.6	7.9	48.7	11.0	133.9	30.0	1.52
F	63.2	11.6	1.24	55.1	10.0	0.51	63.9	9.8	1.40	48.5	8.5	48.8	11.7	146.2	25.0	1.60
M&F	62.8	12.6	1.16	54.6	10.0	0.46	62.5	12.4	1.14	47.0	8.1	48.7	10.5	140.0	20.2	1.45
Normative group																
M	50.0	10.0		50.0	10.0		50.0	10.0		50.0	10.0	50.0	10.0	97.5	22.3	
F	50.0	10.0		50.0	10.0		50.0	10.0		50.0	10.0	50.0	10.0	108.0	22.2	
M&F	50.0	10.0		50.0	10.0		50.0	10.0		50.0	10.0	50.0	10.0	103.1	23.5	

Note. Effect sizes (Cohen's *d*) are based on comparisons with control participants (the PSI normative group). AI = Alienation; Sn = Social Nonconformity; Di = Discomfort; Ex = Expression; De = Defensiveness; M = male; F = female.

Table 4
Accuracy Data in Distinguishing Three Target Groups From Control Participants, for Cut Scores Setting Specificities Close to .80 and Base Rate (for PPP/NPP) of .20

Measure and gender	AUC	Cut score	Predictive power		Sensitivity	Specificity
			PPP	NPP		
Major psychiatric disorder						
AI scale						
M	.912	≥8	.501	.962	.876	.782
F	.898	≥8	.481	.964	.888	.760
M&F	.905	≥8	.491	.963	.882	.771
Significant antisocial behavior						
Sn scale						
M	.846	≥12	.466	.934	.794	.772
F	.827	≥10	.520	.916	.692	.840
M&F	.811	≥11	.462	.918	.717	.791
General psychological distress						
Di scale						
M	.743	≥13	.449	.900	.645	.800
F	.816	≥16	.425	.919	.719	.807
M&F	.773	≥15	.458	.900	.640	.811
General psychological distress						
OSIR						
M	.833	≥115	.476	.917	.706	.806
F	.867	≥127	.494	.942	.803	.794
M&F	.842	≥122	.472	.924	.739	.793

Note. AUC = area underneath the receiver operating characteristic curve; PPP = positive predictive power; NPP = negative predictive power; AI = Alienation; Sn = Social Nonconformity; Di = Discomfort; OSIR = Overall Screening Index-Revised; M = male; F = female.

PPP and NPP. Table 4 presents the PPP–NPP findings. As stated earlier, the cut scores for all PPP–NPP analyses were chosen to provide a specificity as close as possible to .80. The cut score used was a raw score of 9 and above for both men and women separately and for the combined group. For the combined group, the PPP was .491 and the NPP was .963. For men alone, the figures were .501 and .962, respectively; for women, .481 and .964, respectively. Thus, for example, the likelihood that a man scoring at or above the cut score would have a major psychiatric disorder was .501, based on the parameters chosen in this analysis.

Sensitivity. Table 4 shows that the proportions of psychiatric inpatients correctly identified by a cut score of 8 and above on AI were .876 for men, .888 for women, and .882 for the combined group. Corresponding proportions for the control group (specificities) were .782, .760,

and .771, respectively. These data are very similar to the initial validation data reported in the *PSI Manual* (Lanyon, 1973, 1978).

Significant Antisocial Behavior (Sn Scale)

Effect size. Effect sizes for the Sn scale in distinguishing persons with significant antisocial behavior from the control group are shown in Table 3. These effect sizes were very large (approximately 1.4). The effect sizes for the Sn scale in distinguishing the other target groups were small (average of about 0.4).

ROC analysis. As reported in Table 4, the AUC representing the ability of the Sn scale to distinguish between the control group and persons with significant antisocial behavior across a broad range of cut scores and base rates was .811 (95% confidence interval = .786–.836) for the combined group. For men and women sep-

arately, the AUCs were .846 and .827, respectively. Using the criteria cited earlier, these figures indicate excellent discrimination.

PPP and NPP. The cut score used for men was 12 and above; for women, 10 and above; and for the combined group, 11 and above. Table 4 shows that for men the PPP was .466 and the NPP was .934. For women, these figures were .520 and .916, respectively; for the combined group, .462 and .918, respectively.

Sensitivity. Using the same cut scores as for the PPP–NPP analysis, Table 4 shows that the proportions of persons with significant antisocial behavior correctly identified by the Sn scale were .794 for men, .692 for women, and .717 for the combined group. Corresponding proportions for control participants (specificities) were .772, .840, and .791, respectively. As with the AI scale for distinguishing major psychiatric disorder, these results are very similar to the initial validation data reported in the *PSI Manual* (Lanyon, 1973, 1978).

General Psychological distress (Di Scale)

Effect size. Effect sizes for the Di scale in distinguishing persons with general psychological distress from control participants are shown in Table 3. The Di scale also showed moderate to large effect sizes in distinguishing the other two target groups from control participants.

ROC analysis. As reported in Table 4, the AUC representing the ability of the Di scale to distinguish between control participants and persons with general psychological distress across a broad range of cut scores and base rates was .773 (95% confidence interval = .748–.797) for all participants. For men and women separately, the AUCs were .743 and .816, respectively. Using the criteria cited above, these figures indicate acceptable to excellent discrimination.

PPP and NPP. The cut score used for men was 13 and above; for women, 16 and above; and for men and women combined, 15 and above. Table 4 shows that for men the PPP was .449 and the NPP was .900. For women, these figures were .425 and .919, respectively; and for men and women combined, .458 and .900, respectively.

Sensitivity. Using the same cut scores as for the PPP–NPP analysis, Table 4 shows that the proportions of persons with general psycholog-

ical distress correctly identified by the Di scale were .645 for men, .719 for women, and .640 for the combined group. Corresponding proportions for normal participants (specificities) were .800, .807, and .811, respectively.

General Psychological Distress (OSIR)

Effect size. The same analyses as used with each of the three screening scales were conducted for persons with general psychological distress with the OSIR. These results are shown in Tables 3 and 4. As expected, the OSIR showed very large effect sizes. It also showed very large effect sizes for psychiatric inpatients and smaller but still large effect sizes for persons with significant antisocial behavior.

ROC analysis. The AUC representing the ability of the OSIR to distinguish between the control group and persons with general psychological distress across a broad range of cut scores and base rates was .842 (95% confidence interval = .822–.863) for all participants. For men and women separately, the AUCs were .833 and .867, respectively. Using the criteria cited earlier, these figures indicate excellent discrimination. Comparison with the AUCs based on Di scores alone showed better discrimination ($p < .05$) for the OSIR for men alone and for men and women combined.

PPP and NPP. The cut score used for men was 115 and above; for women, 127 and above; and for men and women combined, 122 and above. Table 4 shows that for men the PPP was .476 and the NPP was .917. For women, these figures were .494 and .942, respectively; and for the combined group, .472 and .924. These figures are numerically higher than for the Di scale alone, with an average increase of about four percentage points in PPP and about two percentage points in NPP.

Sensitivity. Using the same cut scores as for the PPP–NPP analysis, Table 4 shows that the proportions of persons with general psychological distress correctly identified by the OSIR were .706 for men, .803 for women, and .739 for the combined group. Corresponding proportions for normal participants (specificities) were .806, .794, and .793, respectively. Once again, these figures are numerically higher than the results based on the Di scale alone.

Further Analyses

The ability of the OSIR to discriminate the control group from the remaining two target groups was also studied using ROC analyses. ROCs for participants with major psychiatric disorder when compared with controls were .899 for men, .845 for women, and .873 for the combined group. These figures are comparable to the accuracy achieved by the AI scale alone (see Table 4). For participants with significant antisocial behavior, the ROCs were .758 for men, .720 for women, and .740 for the combined group. All of these figures are significantly lower ($p < .05$) than the accuracy achieved by the Sn scale alone.

Discussion

The hypotheses of the study were supported. On the basis of the ROC and effect size data, the three screening scales showed their strongest discriminations in identifying the target group for which they were designed, and did so at an acceptable level of accuracy. The Di scale was the least strong in its task of identifying persons with general psychological distress; however, the OSIR, developed empirically for the same purpose, showed stronger discrimination. Examination of the mean scores and effect size data in Table 3 shows that elevations on the AI scale (developed to screen for persons with major psychiatric disorder) were also common among participants with significant antisocial behavior and general psychological distress, a finding that would not be unexpected as some of these persons would be expected to have major psychiatric disorder. It is also noted that male psychiatric inpatients showed higher levels of general psychological distress than female psychiatric inpatients, as indicated by their relative elevations on Di and the associated effect sizes. This finding suggests that women might have been more readily hospitalized than men, although additional data would be needed to study this hypothesis.

In selecting cut scores for the PPP–NPP and sensitivity–specificity analyses, raw scores were selected that corresponded to a *T* score close to 60. This was the guideline originally proposed and used in initial validation studies (Lanyon, 1973), and although selection of cut scores should be based on the needs of each

individual situation in which the PSI is used, the present results suggest that a *T* score of around 60 continues to be an appropriate first approximation to be considered.

To compare this study's results for the OSIR with Overall's (1974) original work, the cut scores that yielded the optimal discrimination in this study for persons with general psychological distress were determined from the percentile distributions. These data showed optimal discriminations of 76.4% for men, 80.0% for women, and 76.7% for the combined group. These results can be compared with the figure of 82.5% reported by Overall, who found the same optimal cut score for both men and women. Some shrinkage would be expected on cross-validation, particularly because of the small size of Overall's original group ($N = 126$). The AUC data reported in Table 4 slightly exceeded Overall's optimal hit rate. It is noted that the development of the OSIR involved the same control group as used in the present comparisons. Thus, the present analyses cannot be considered an independent cross-validation.

The best discrimination of all was shown by the AI scale for major psychiatric disorder. This is a reasonable finding, as the definition of major psychiatric disorder is somewhat narrower than the other two target groups. Despite the limitations of the OSIR findings indicated earlier, the findings for this index were consistent with the notion that general psychological distress encompasses major psychiatric disorder.

To examine the possibility that Di scale scores are particularly prone to distortion through defensiveness, a further set of ROC analyses was conducted in discriminating persons with general psychological distress from the control group, using a simple corrected score: the raw-score sum of the Di scale and the De scale. For men alone, the AUC increased from .743 to .786, with comparable small increases in PPP–NPP and sensitivity. For women, the accuracy figures were essentially unchanged. These analyses were repeated, using a correction of half the De raw score, and this analysis resulted in a smaller increment for men and (once again) no increment for women. Thus, the use of a simple correction for defensiveness did not improve discrimination as much as using the OSIR (which also includes a fraction of the De scale in addition to a fraction of the AI scale).

In light of the view that equal weighting of predictors is at least as effective as differential weights if the predictors are substantially correlated (Dawes, 1979), the OSIR-related analyses were repeated using the unweighted sum of AI, Sn, and De as the predictor. These analyses resulted in accuracy figures that were in general slightly numerically lower than for the weighted index of the OSIR.

It is noted that the definition of significant antisocial behavior was rather broad, including individuals with a wide variety of crimes plus state hospital patients diagnosed with antisocial personality disorder (see Table 1). A small group of the incarcerated individuals, all male ($n = 16$), had been accused of homicide and were facing the death penalty. These participants could be considered the most "severe" of the total group, and comparison data supported this view. Thus, their mean T score on the Sn scale was 68.4 (compared with 65.1 for the complete group of men), and ROC analysis showed an AUC of .861, compared with .846 for the complete group of men.

It was wondered whether anything could be learned about persons with significant antisocial behavior who scored low on the Sn scale and therefore were not correctly identified by Sn. Visual inspection of the data for those individuals showed that some of them fell into one of two groups: high scorers on the measures of psychological difficulty (AI, Di, or both) or high scorers on defensiveness (De). There is thus the possibility that a stepwise procedure involving two or more scales could improve the discrimination. Such a possibility would be consistent with my experience that a small proportion of incarcerated persons, especially those awaiting trial, have significant psychological dysfunction rather than significant antisocial characteristics.

For the Sn and Di scales, the slight differences in norms for men and women (see Lanyon, 1973, 1978) suggested that it could be advantageous in the present study to consider men and women separately. Related is the present finding that some target characteristics were slightly more readily identified for one gender than for the other; for example, general psychological distress for women and perhaps significant antisocial behavior for men. This is an additional reason for considering men and women separately when using these data in an objective manner.

In regard to education level, the two target groups for whom education data were available (persons with major psychiatric disorder and persons with significant antisocial behavior) showed a mean education level 2–3 years lower than the control participants (see Table 2). However, inspection of the *PSI Manual* (Lanyon, 1973, 1978) shows that this was also the case for the participants on whom the scales were originally developed and validated. It is also noted that psychiatric disorder in general has a small negative relationship to socioeconomic level (e.g., Dohrenwend et al., 1992). Thus, the difference in education level probably does not interfere with the validity of the discriminations that can be made. Nevertheless, caution should be observed when interpreting the scores of participants with very low educational levels (in addition to considering the question of adequate reading ability).

Comparing the sensitivity–specificity rates achieved in the present study with the pilot data reported in the *PSI Manual* (Lanyon, 1973, 1978) in making the same discriminations shows comparable accuracy for the AI scale in identifying psychiatric inpatient status and slightly lower accuracy for the Sn scale in identifying men with significant antisocial behavior. Examination of the mean Sn scale scores of the subgroups of participants listed in Table 1 shows somewhat lower means for the state hospital group (with a diagnosis of antisocial personality disorder) than for incarcerated persons, with the former accounting for about 30% of the total group. Reanalysis omitting the state hospital group showed improved accuracy rates, very close to those originally reported in the *PSI Manual*.

How do the current results compare with those for other screening tests? The only relevant comparison data that could be found involved the Millon Clinical Multiaxial Inventory—III (Millon, Davis, & Millon, 1997). However, the uniqueness of Millon's system of converting raw scores to scaled scores (base rate scores, in his terminology), plus other factors, including the larger number of Millon's diagnostic categories, variability in base rates (prevalence, in his terminology), lack of information regarding participant groups and the origins of the data, and ambiguity as to the various cut scores used, make direct comparison with his work of uncertain value (see Hsu, 2002, for a

discussion of additional issues). With caution due to these limitations, examination of the data for the Millon Clinical Multiaxial Inventory—III diagnostic categories that are the most relevant to the three participant groups in the present study shows equal or better accuracy for the results involving the PSI.

The question can be raised as to whether the PSI screening scales provide an increment in accuracy over even shorter screening procedures—in the extreme, single items such as (for significant antisocial behavior) “I have never been in trouble with the law.” Although such a study has not been conducted, shorter scales are less reliable. Note is made of Jackson’s (1974, 1976) decision to increase the number of items on his personality scales from 16 for the Personality Research Form to 20 for the Jackson Personality Inventory in order to provide adequate reliability.

The present study has several limitations. First, in regard to the possible effect of ethnicity on PSI scores, no relevant data were available for the control group. However, such data were available for 98 participants who had undergone forensic evaluation (some had also been included as participants in the target group with significant antisocial behavior). Of these individuals, 78 were Caucasian and 20 were Hispanic ($n = 14$), African American ($n = 2$), or Native American ($n = 4$). Gender was comparably represented. The mean scores for the two groups did not differ by more than one raw-score point on any of the five PSI scales. Although these data are inadequate for drawing conclusions about the effect of ethnicity, they do suggest that such differences are probably not large. However, more extensive data are needed.

Studies bearing on the issue of ethnicity-related differences on the Minnesota Multiphasic Personality Inventory–2 (MMPI-2) were recently reviewed by Greene, Robin, Albaugh, Caldwell, and Goldman (2003), who also reported their own extensive study comparing Caucasians and American Indians. They concluded that any differences appeared to reflect genuine differences between the groups tested rather than test bias. Similarly, a meta-analysis of 25 MMPI/MMPI-2 studies by Hall, Bansal, and Lopez (1999) comparing Latino Americans, African Americans, and European Americans “did not suggest substantive differences from

either a statistical or clinical perspective” (p. 186).

Another limitation of the present study involves the use of PSI control-group norms that were developed in the early 1970s as a comparison group for contemporary purposes. This issue was addressed in part by assessing the comparability of the original norms for ages 18–25 ($N = 274$) with an available normative group of the same age range, education, and gender composition ($N = 103$). The mean scores for the two groups did not differ by more than one raw-score point on any of the five PSI scales. Again, although these data are obviously not definitive, they suggest some degree of comparability between the old norms and likely contemporary norms. Once again, further data are needed.

Utility in the Public Sector

Elsewhere, I have addressed the importance of developing and propagating procedures for mental health–related assessment that are maximally cost effective in terms of benefit to the public (Lanyon, 1971, 1972). Two points made at that time continue to be applicable today. First, whenever the available decision options are limited and the maximum attainable accuracy is not high, simple assessment procedures tend to be more cost effective than more complex procedures. Second, the issue of costs is a complex one and must include factors such as productive work time lost as a result of problems not treated, plus a variety of less tangible and difficult-to-define social costs.

Much of the test information represented in Tables 1–3 was provided in the context of routine screening relevant to mental health in public practice settings. In some, such as prison settings, the task was to economically identify persons in their general population who were at risk for mental health or other behavioral problems. In others, such as community mental health settings, the screening task was undertaken to narrow the focus of a subsequent, more comprehensive evaluation.

Effectiveness data have been collected in several pilot situations that exemplify these uses. In one, the PSI was administered to 52 patients before they underwent full-mouth extractions at a large dental clinic. Adverse psychological reactions had been found to be

common in this setting. Eight protocols were put aside owing to excessively defensive responding; for the remainder, a *T* score of 64 or more on the Al, Sn, or Di scale was designated as the criterion for potential psychological problems that might interfere with treatment. Nine such persons were identified. A further 9 patients were selected at random from the remaining pool, and these 18 individuals were then independently judged as “difficult patient” or “not difficult” based on subsequent file information. The PSI and the independent case judgments were in agreement for 16 of the possible 18 instances.

In another use, the PSI was administered to an entire freshman class of 135 student nurses. After 6 months, faculty advisors who had worked with the students made a joint decision regarding serious adjustment difficulties of a psychiatric nature. Six students were designated as such. Of 6 students who had scored 65 or more on the Al scale, 3 were among those identified by the faculty, another had already withdrawn, and another had serious difficulty with the English language.

The results of these pilot studies, although obviously not definitive, serve to illustrate some appropriate uses of the PSI. Together with the large-scale data on accuracy presented in Table 3, they support the continued use of this instrument as an inexpensive method of screening for mental health problems.

In conclusion, the present cross-validation data indicated that the Al and Sn scales and the OSIR showed good discriminations for persons with major psychiatric disorder, significant antisocial behavior, and general psychological distress, respectively. The Di scale also performed satisfactorily; however, the OSIR outperformed the Di scale in identifying persons with general psychological distress and also performed well in identifying psychiatric inpatients. Adding a simple correction for defensiveness to the Di scale did not improve its success to the level shown by the OSIR. Although it is possible that other weighted combinations of scales could improve on the success rates reported here, the purpose of the present study was to comprehensively evaluate the success of the three PSI screening scales for the tasks they were designed to do.

References

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist, 34*, 571–582.
- Dohrenwend, B. P., Levav, I., Shrout, P. E., Schwartz, S., Naveh, G., Link, B., Skodol, A., & Stueve, A. (1992). Socioeconomic status and psychiatric disorders: The causation-selection issue. *Science, 255*, 946–952.
- Greene, R. L., Robin, R. W., Albaugh, B., Caldwell, A., & Goldman, D. (2003). Use of the MMPI-2 in American Indians: II. Empirical correlates. *Psychological Assessment, 15*, 360–369.
- Hall, C. G. N., Bansal, A., & Lopez, I. R. (1999). Ethnicity and psychopathology: A meta-analytic review of 31 years of comparative MMPI–MMPI-2 research. *Psychological Assessment, 11*, 186–197.
- Hays, W. L. (1994). *Statistics* (5th ed.). Fort Worth, TX: Harcourt Brace.
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York: Wiley.
- Hsu, L. M. (2002). Diagnostic validity studies and the MCMI-III. *Psychological Assessment, 13*, 410–422.
- Jackson, D. N. (1974). *Personality Research Form: Manual* (Rev. ed.). Port Huron, MI: Research Psychologists Press.
- Jackson, D. N. (1976). *Jackson Personality Inventory: Manual*. Port Huron, MI: Research Psychologists Press/Sigma Assessment Systems.
- Lanyon, R. I. (1970). Development and validation of a psychological screening inventory. *Journal of Consulting and Clinical Psychology, 35*(1, Pt. 2), 1–24.
- Lanyon, R. I. (1971). Mental health technology. *American Psychologist, 26*, 1071–1076.
- Lanyon, R. I. (1972). A technological approach to the improvement of decision making in mental health services. *Journal of Consulting and Clinical Psychology, 39*, 43–48.
- Lanyon, R. I. (1973). *Psychological Screening Inventory: Manual*. Goshen, NY: Research Psychologists Press.
- Lanyon, R. I. (1978). *Psychological Screening Inventory manual* (2nd ed.). Port Huron, MI: Research Psychologists Press.
- Lanyon, R. I. (1993). Development of scales to assess specific deception strategies on the Psychological Screening Inventory. *Psychological Assessment, 5*, 342–329.
- Lanyon, R. I. (2005). Construct validity of the misrepresentation scales of the Psychological Screen-

- ing Inventory. *Journal of Personality Assessment*, 85, 197–206.
- Millon, T., Davis, R., & Millon, C. (1997). *Millon Clinical Multiaxial Inventory-III: Manual* (2nd ed.). Minneapolis, MN: NCS Pearson.
- Overall, J. (1974). Validity of the Psychological Screening Inventory for psychiatric screening. *Journal of Consulting and Clinical Psychology*, 42, 717–719.
- Rice, M. E., & Harris, G. T. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology*, 63, 737–748.
- Swets, J. A. (1992). The science of choosing right decisions in high-stakes diagnostics. *American Psychologist*, 47, 522–532.

Received October 20, 2005

Revision received February 24, 2006

Accepted March 28, 2006 ■

E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at <http://watson.apa.org/notify/> and you will be notified by e-mail when issues of interest to you become available!