Psychological Assessment
2002, Vol. 14, No. 4, 401–409

1040-3590/02/$5.00 DOI: 10.1037//1040-3590.14.4.401

Incremental Validity of Selected MMPI–A Content Scales in an Inpatient Setting

Robert E. McGrath
Fairleigh Dickinson University

David L. Pogge
Four Winds Hospital

John M. Stokes
Pace University

To date, relatively few studies have been published evaluating the validity or incremental validity of the content scales from the adolescent version of the Minnesota Multiphasic Personality Inventory (MMPI–A; J. N. Butcher et al., 1992). A sample of 629 psychiatric inpatient adolescents who had completed the MMPI–A was used to evaluate the ability of selected clinical and content scales to predict conceptually related clinical variables. Criteria were based on clinician ratings, admission and discharge diagnoses, and chart reviews. Results from hierarchical multiple and logistic regression analyses indicated the content scales offered incremental validity over the clinical scales and supported the use of the content scales as an adjunct to the traditional clinical scales.

In most cases, the original clinical scales of the Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1983) comprised those items that significantly discriminated between members of a diagnostic group and the normative sample. Over time, this approach to item selection has become controversial. The clinical scale items are extremely heterogeneous in content, complicating their interpretation and potentially attenuating their predictive ability. Furthermore, many of the clinical scale items are not clearly relevant to the measurement of psychopathology. A number of authors have questioned whether these so-called "subtle items" add to the clinical scales in a substantive way (Jackson, 1971; Weed, Ben-Porath, & Butcher, 1990), though others have been more optimistic about their contribution (e.g., see Hollrah, Schlottmann, Scott, & Brunetti, 1995). The extensive item overlap between the clinical scales has also been criticized.

In response to concerns raised about the multidimensionality of the clinical scales, Wiggins (1966) undertook an extensive reorganization of the entire MMPI item pool based on item content. Using both rational and statistical criteria, he developed 13 content-based scales that demonstrated less dimensional complexity and item overlap than the clinical scales.

During the development of the revised version of the MMPI (MMPI–2; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989), Butcher, Graham, Williams, and Ben-Porath (1990) concluded that the content of the MMPI item pool had changed sufficiently to warrant generating a new set of content scales. On the basis of rational and statistical methods similar to those used by Wiggins (1966), they defined 15 content scales that are homogeneous in content and demonstrate minimal item overlap.

Several studies have since demonstrated the incremental validity of the MMPI–2 content scales when combined with the traditional clinical scales. Ben-Porath, Butcher, and Graham (1991) found the Depression and Bizarre Mentation content scales improved the prediction of depression and schizophrenia in 160 inpatients over the standard clinical and validity scales, with a median improvement in the proportion of variance predicted of 8%. In contrast, none of the clinical and validity scales that they examined significantly improved the prediction of diagnosis over the content scales.

Ben-Porath, McCully, and Almagor (1993) found the MMPI–2 content scales significantly improved the prediction of conceptually related self-report scales in 596 college students, with a median improvement in the proportion of variance predicted of 10%. In contrast to the earlier study, clinical scales also significantly improved the prediction of criterion scores over the content scales, although the median improvement in the proportion of variance predicted was only 3%.

Archer, Aiduk, Griffin, and Elkins (1996) used both clinician ratings and self-report scores as criteria in a sample of 597 psychiatric inpatients. The median improvement in the proportion of

Robert E. McGrath, School of Psychology, Fairleigh Dickinson University; David L. Pogge, Psychology Service, Four Winds Hospital, Katonah, New York; John M. Stokes, Department of Psychology, Pace University.

This project was supported in part by a grant from the Fairleigh Dickinson University Grant-in-Aid Program. We thank Ana Cragnolino, Judy Hayman, Teresa Piacentini, Doug Wayland-Smith, and Michele Zaccario for their invaluable assistance in gathering the data used for this study. We also gratefully acknowledge Samuel Klagsbrun, Martin Buccolo, Janet Segal, and the administration of Four Winds Hospital, Katonah, New York, for their support in the collection of data for this article. The authors are solely responsible for its content.

Correspondence concerning this article should be addressed to Robert E. McGrath, School of Psychology T-WH1-01, Fairleigh Dickinson University, Teaneck, New Jersey 07666. E-mail: mcgrath@fdu.edu

variance predicted when the content scales were added to the clinical scales was 3%. Adding the clinical scales to the content scales resulted in similar improvements in concurrent validity.

In the most recent study evaluating the incremental validity of the MMPI–2 content scales, Barthlow, Graham, Ben-Porath, and McNulty (1999) found both clinical and content scales enhanced the prediction of conceptually related clinician ratings in a sample of 699 psychiatric outpatients. The median improvement in the prediction of clinician ratings by both the content and clinical scales was 2%.

Content scales were also developed for the adolescent version of the MMPI (MMPI–A; Butcher et al., 1992). Williams, Butcher, Ben-Porath, and Graham (1992) again used both rational and statistical criteria to generate 15 MMPI–A content scales, 11 of which are essentially equivalent to content scales from the MMPI–2.

Several studies have been published examining the validity of the MMPI–A content scales. Cashel, Rogers, Sewell, and Holliman (1998) identified clinical correlates of content scales in 99 delinquent adolescents, and Pena, Megargee, and Brody (1996) demonstrated the validity of the scales in 162 male delinquents. Arita and Baer (1998) demonstrated convergent and discriminant validity for certain MMPI–A content scales using other self-report measures as criteria.

To date, only two studies have been published evaluating the incremental validity of the MMPI–A content scales. Kopper, Osman, Osman, and Hoffman (1998) examined the incremental validity of the content scales for the prediction of suicide risk in 143 inpatient adolescents. The study was problematic in several ways. Score on the Suicide Probability Scale (Cull & Gill, 1988), a self-report measure, was the only criterion used. The authors combined only one clinical scale with eight content scales in each of their analyses. Finally, incremental validity was reported only for those content scales that significantly enhanced prediction. Across six analyses examining the incremental validity of the content scales over the clinical scales, the median increase in the proportion of variance predicted was .33.

Archer and Krishnamurthy (1997) found that the content scales enhanced the prediction of depressive and conduct disorder diagnoses in 152 adolescents from a variety of settings. However, the authors used stepwise rather than hierarchical analyses that included both MMPI–A and Rorschach variables, so it is not possible to determine the increment in validity that would have resulted from adding content scales directly to the standard clinical scales.

This literature review indicates there is relatively little information available concerning the incremental validity of the MMPI–A content scales when added to the clinical scales. The two studies that have been published on the topic addressed a limited set of criterion variables. Furthermore, neither indicated the increment in validity provided by the full set of content scales entered into the analyses, as is typically considered appropriate when conducting incremental validity tests. Finally, both adult and adolescent studies have relied heavily on the use of other self-report measures as criteria. This is a questionable basis for establishing the incremental validity of the content scales, because the outcome could be largely determined by the degree of overlap in item content between the content scale, the clinical scale, and the criterion scale.

In the current study, we examine the incremental validity of the MMPI–A content scales when combined with the traditional clinical scales. Hierarchical analyses were used to estimate the increment in concurrent validity offered by the content scales. The criteria were limited to variables based solely on clinical judgments to avoid possible method artifacts associated with the use of self-report criteria.

Method

Participants

The initial sample consisted of 913 adolescents who completed the MMPI–A on admission to Four Winds Hospital, a private psychiatric facility in the New York metropolitan area. None of the youths in the sample omitted more than 18 items from the MMPI–A. The following criteria for potentially invalid protocols eliminated 284 youths from the sample: Lie (L) Scale $>$ 65 *T* (n = 187), Frequency (F) Scale $>$ 89 *T* (n = 8), Correction (K) Scale $>$ 65 *T* (n = 107), Variable Response Inconsistency (VRIN) Scale $>$ 79 *T* (n = 27), or True Response Inconsistency (TRIN) Scale $>$ 79 *T* (n = 27) (Arita & Baer, 1998).

The loss of 31% of cases because of potentially invalid responding seemed high, though it is consistent with the 21% that Arita and Baer (1998) eliminated from their smaller sample using the same criteria. In response to the suggestion that these criteria for L and K might be overly conservative (R. P. Archer, personal communication, March 18, 2001), the analyses were repeated with two alternate sets of cut scores. In the first replication, youths were excluded if K or L $>$ 70 *T*, or if both K and L $>$ 65 *T*. With these criteria, the number of cases retained for the analyses increased, but the mean correlation between the MMPI scales and the criteria declined, resulting in little difference in the outcomes. In the second replication, K and L were not used to exclude cases at all. Although this modification dramatically increased the number of cases retained for the analyses (only 54 cases were excluded), the number of significant outcomes actually declined because of the reduced size of the correlations. Accordingly, the alternative criteria were rejected as a basis for excluding potentially invalid protocols. Table 1 provides demographic data for the final sample of 629.

Table 2 provides descriptive statistics for each of the standard MMPI–A validity, clinical, and content scales. Most of the sample means fell within 1 standard deviation of the normative sample means. This finding is not surprising, given that previous research has consistently indicated mean scores on both the MMPI–A clinical and content scales are within normal limits across populations (see Archer, Handel, & Lynch, 2001, for a review), suggesting the current sample is typical of adolescents in psychiatric settings. However, this finding potentially suggests the contradictory position to that explored in the previous paragraph, that the traditional criteria for underreporting are too liberal for the MMPI–A. Analyses were completed again, this time tightening the criteria for underreporting. In the first replication, cases were excluded if either L or K exceeded 60. These criteria excluded 400 cases (44% of the sample), with no increase in mean correlations when compared with the original validity criteria. The final replication excluded all cases where L or K exceeded 65. These criteria ruled out 550 cases, yet the mean correlation was actually lower than it was with the original criteria. On the basis of these findings, it was concluded that Arita and Baer's (1998) criteria for determining invalid protocols are at least as good as any other set of criteria evaluated in this study.

Procedure

The MMPI–A is part of the standard admission battery for the facility in which the study was conducted. Participants were readministered the inventory if they omitted more than 30 items.

Table 1
Demographic Characteristics of the Sample

Characteristic	*M*		*SD*	
	Boys	Girls	Boys	Girls
Age (years)	14.9	14.8	1.3	1.1
Grade level	9.2	9.3	1.4	1.2
Admission Axis V	31.7	31.3	7.6	7.6
Ethnicity (%)				
White	71.0	68.6		
Black	13.9	14.9		
Hispanic	12.5	11.8		
Other	2.6	4.7		
Learning disability (%)	22.7	10.8		
Admission diagnoses (%)				
Psychosis	15.1	15.1		
Conduct disorder	24.7	20.9		
Depression	64.5	72.9		
Discharge diagnoses (%)				
Psychosis	18.1	18.5		
Conduct disorder	26.6	21.2		
Depression	59.5	65.8		

Note. For boys, $n = 304$; for girls, $n = 325$.

Three sources of clinician-based criterion variables were available for the present study. First, each youth's primary therapist was asked to complete the Hopkins Psychiatric Rating Scale (HPRS; Derogatis, 1983). The HPRS is a rating scale similar in wording and format to the Brief Psychiatric Rating Scale (Overall & Gorham, 1962). It was developed to parallel the Symptom Checklist 90 (Derogatis, 1983). It includes one item representing each of the nine symptom dimensions from that inventory, as well as eight additional items representing specific symptom dimensions and one global severity item.

A number of the HPRS items require the clinician to consider multiple issues when making the ratings. To generate more homogeneous criteria for the study, we used principal-components analysis of the 17 HPRS symptom items to estimate latent structures underlying the ratings. The first two unrotated components were each considered conceptually related to one clinical and one content scale. The first component seemed to be a measure of negative affect or general distress, loading .30 or higher on all 17 items. Items reflecting depression were the primary contributors to the second component. This component loaded .20 or higher on items tapping depression, abjection–disinterest, motor retardation, and sleep disturbance, and loaded −.30 or lower on items reflecting excitement, euphoria, and hostility. Scores for these two components were generated for 576 youths in the sample.

Second, diagnostic data were available for every member of the sample. The clinical team was allowed to identify up to five Axis I diagnoses for each youth at both admission and discharge. On the basis of these diagnoses, six variables were created based on whether the youth received a psychotic diagnosis, a depressive diagnosis, or a conduct disordered diagnosis at admission or discharge.

The third source of clinician data was a chart review completed on each youth in the sample. Doctoral students in clinical psychology completed the chart reviews. The form used for this review was based on an earlier one introduced by Williams and Butcher (1989) during the development of the MMPI–A. Based on a combination of conceptual considerations and correlations between chart review items, five clusters of items were identified that were thought to be conceptually related to at least one clinical and content scale. The self-harm cluster consisted of four items having to do with suicidal ideation and with a history of self-mutilation and suicide attempts. The conduct cluster consisted of eight items suggesting problems with anger, criminal behavior, and oppositionalism. The bizarreness cluster consisted of five items having to do with chart data indicating bizarre thinking and sensory experiences as well as paranoia. The obsessiveness cluster consisted of four items suggesting compulsive or obsessive tendencies. The depression cluster consisted of five items indicating traditional symptoms of depression. Table 3 lists the items included in each cluster.

Given the archival nature of the data, interrater reliability data were unavailable for the clinical judgments that provided the basis for the criterion variables. Chart reviews were completed twice for 100 participants in the original sample. One of the items included in the five clusters was quantitative (the number of previous suicide attempts). The intraclass correlation between raters for this item was .92. The remaining 25 cluster items were categorical. Kappa coefficients for these items ranged between .53 and 1.00, with a mean of .74 and a median of .75.

Table 4 provides the list of criterion variables as well as the clinical and content scales that were considered conceptually related to each. The number of clinical and content scales used as predictors for each criterion was equated.

Table 2
Mean Profiles and Standard Deviations for the Sample

Scale	*M*		*SD*	
	Boys	Girls	Boys	Girls
Validity scales				
L	50.6	51.3	7.1	7.8
F1	59.2	63.5	11.8	13.7
F2	54.8	54.5	11.7	10.3
K	49.6	49.7	8.6	7.8
VRIN	51.8	52.3	9.0	8.8
TRIN	58.3	58.3	6.6	6.2
Clinical scales				
Scale 1: Hypochondriasis	53.4	55.3	12.2	12.1
Scale 2: Depression	58.4	61.5	11.7	12.6
Scale 3: Hysteria	54.5	57.4	9.7	13.3
Scale 4: Psychopathic Deviate	62.9	62.4	12.3	11.5
Scale 5: Masculinity/Femininity	49.2	54.0	10.1	10.4
Scale 6: Paranoia	57.2	57.5	12.4	11.2
Scale 7: Psychasthenia	54.9	55.3	13.4	12.0
Scale 8: Schizophrenia	56.6	57.0	13.4	12.3
Scale 9: Hypomania	55.7	53.9	11.8	11.5
Scale 0: Social Introversion	50.6	51.8	11.1	10.9
Content scales				
Anxiety	55.4	57.5	12.4	12.3
Obsessiveness	52.0	51.4	10.6	10.7
Depression	58.5	59.6	13.5	12.0
Health Concerns	53.5	54.8	11.4	11.0
Bizarre Mentation	53.2	52.5	12.0	11.6
Anger	54.8	55.4	12.0	12.1
Cynicism	52.2	53.3	10.8	9.5
Conduct Problems	56.2	58.1	12.8	12.7
Low Self-Esteem	55.5	55.6	12.9	12.1
Social Discomfort	50.8	51.6	12.2	13.2
Family Problems	59.6	60.1	13.1	13.5
Negative Treatment Indicators	55.0	56.8	14.2	13.4
Alienation	55.7	55.6	12.2	10.5
Low Aspiration	54.3	56.8	11.2	12.7
School Problems	62.4	64.1	14.3	13.5

Note. L = Lie; F1 and F2 = subscales of the Infrequency scale; K = Correction; VRIN = Variable Response Inconsistency; TRIN = True Response Inconsistency.

Table 3
Contents of Chart Review Clusters

Cluster	Item
Self-harm	No. of suicide attempts
	Suicide attempts?
	Suicidal ideation?
	Self-mutilation?
Conduct	Anger outbursts?
	Oppositional behavior?
	Mood swings?
	School suspensions?
	Violent behavior?
	Criminal behavior?
	Homicidal ideation?
	Impulsivity?
Bizarreness	Bizarre behavior?
	Bizarre thoughts?
	Tangential thinking?
	Hallucinations?
	Paranoid ideation?
Obsessiveness	Compulsions?
	Obsessions?
	Phobias?
	Sexually active (–)?
Depression	Depressed mood?
	Lethargy?
	Sleep difficulties?
	Grandiosity (–)?
	Social withdrawal?

Note. Items followed by a question mark are dichotomous; items followed by (–) loaded negatively on the cluster.

Results

Incremental Validity

Hierarchical regression procedures were used to evaluate the incremental validity of the clinical and content scales. The HPRS ratings and chart review cluster scores were quantitative. Hierarchical multiple regression analyses were conducted, and incremental validity tests were based on the differences between the resulting multiple correlations. The diagnostic variables were dichotomous. Hierarchical logistic regression analyses were used instead, and incremental validity tests were based on the difference between the chi-square tests for the more and less restricted models. Given the substantial sample size, tests were conducted separately for boys and girls to evaluate whether gender moderated the incremental validity of the content scales.

The content scales. Table 5 provides results from the content scale incremental validity analyses for boys. Table 6 provides the same information for girls. For diagnostic variables, the correlations provided are McFadden's rho squared, which is not a true correlation coefficient. These values and the differences between these values should not be taken literally as indicators of the proportion of variance predicted (Tabachnik & Fidell, 1996).

Among the boys, 5 out of 13 incremental validity tests were significant, whereas 8 out of 13 analyses were significant for girls. Overall, the content scales did not add much to the prediction of the HPRS component scores. The results were stronger for diagnosis, where half of the tests of incremental validity were significant. In particular, the Bizarre Mentation scale improved the prediction of psychotic diagnoses at admission for both boys and girls.

The content scales were even more effective at improving the prediction of the chart-based clusters, where 7 out of 10 analyses were significant. The content scales improved the prediction of self-harm behaviors, conduct problems, and bizarreness in both boys and girls. Across the 14 analyses excluding the diagnostic criteria, the mean increment in the proportion of variance predicted was .02 (*Mdn* = .01) when the content scales were added to the clinical scales.

The clinical scales. Tables 7 and 8 provide results concerning the incremental validity of the clinical scales over the content scales. Only 3 of 13 analyses were significant among the boys, and 4 were significant for the girls. The pattern of the significant findings is consistent across genders. None of the analyses involving diagnosis were significant, suggesting the content scales are more useful than the clinical scales for predicting the three largest categories of diagnoses in adolescents. The clinical scales did enhance the prediction of chart clusters indicating bizarreness and depression, as well as self-harm in boys and obsessiveness in girls. Overall, the mean increment in the proportion of variance predicted, excluding diagnostic analyses, was .02 (*Mdn* = .01) when the clinical scales were added to the content scales.

Concurrent Validity

Although not the focus of the study, Tables 5 through 8 also provide information about the concurrent validity of the clinical

Table 4
Predicted Relationships Between Minnesota Multiphasic Personality Inventory Scales and Criterion Variables

Criterion variable	Clinical scale	Content scale
HPRS components		
Negative affect	7	Anx
Depression	2	Dep
Diagnosis		
Admission psychotic	8	Biz
Admission depression	2	Dep
Admission conduct disorder	4, 9	Ang, Con
Discharge psychotic	8	Biz
Discharge depression	2	Dep
Discharge conduct disorder	4, 9	Ang, Con
Chart review clusters		
Self-harm	2	Dep
Conduct	4, 9	Ang, Con
Bizarreness	8	Biz
Obsessiveness	7	Obs
Depression	2	Dep

Note. HPRS = Hopkins Psychiatric Rating Scale; 7 = Scale 7: Psychasthenia; Anx = Anxiety; 2 = Scale 2: Depression; Dep = Depression; 8 = Scale 8: Schizophrenia; Biz = Bizarre Mentation; 4 = Scale 4: Psychopathic Deviate; 9 = Scale 9: Hypomania; Ang = Anger; Con = Conduct Problems; Obs = Obsessiveness.

Table 5
Incremental Validity for the Content Scales in Boys

Criterion	Predictor	R^2	F or χ^2	*df*	Change analysis R^2	Change analysis F or χ^2	Change analysis *df*
HPRS components							
Negative affect	7	.03	7.2*	1, 285			
	Anx	.03	3.6*	2, 284	.00	0.00	1, 284
Depression	2	.03	8.7*	1, 285			
	Dep	.03	4.4*	2, 284	.00	0.00	1, 284
Diagnosis							
Admission							
Psychotic	8	.04	10.3*	1			
	Biz	.06	15.6*	2	.02	5.36*	1
Depression	2	.01	5.2*	1			
	Dep	.02	6.0*	2	.01	0.84	1
Conduct disorder	4, 9	.00	0.9	2			
	Con, Ang	.01	4.8	4	.01	3.84	2
Discharge							
Psychotic	8	.02	5.1*	1			
	Biz	.03	7.4*	2	.01	2.25	1
Depression	2	.00	0.6	1			
	Dep	.01	2.1	2	.01	1.53	1
Conduct disorder	4, 9	.01	4.9	2			
	Con[a], Ang	.03	11.5*	4	.02	6.67*	2
Chart clusters							
Self-harm	2	.13	45.8*	1, 301			
	Dep	.17	30.3*	2, 300	.04	12.98*	1, 300
Conduct	4[a], 9	.03	4.6*	2, 299			
	Con, Ang[a]	.07	5.9*	4, 297	.04	7.06*	2, 297
Bizarreness	8	.16	56.7*	1, 302			
	Biz	.18	33.6*	2, 301	.02	9.21*	1, 301
Obsessiveness	7	.01	2.0	1, 301			
	Obs	.01	1.0	2, 300	.00	0.00	1, 300
Depression	2	.14	47.2*	1, 301			
	Dep	.14	24.0*	2, 300	.00	1.04	1, 300

Note. If two numbers are listed for *df*, the previous statistic is an *F* value; if one number is listed for *df*, the previous statistic is a chi-square value. R^2 refers to McFadden's rho squared. HPRS = Hopkins Psychiatric Rating Scale; 7 = Scale 7: Psychasthenia; Anx = Anxiety; 2 = Scale 2: Depression; Dep = Depression; 8 = Scale 8: Schizophrenia; Biz = Bizarre Mentation; 4 = Scale 4: Psychopathic Deviate; 9 = Scale 9: Hypomania; Con = Conduct Problems; Ang = Anger; Obs = Obsessiveness.
[a] Indicates a predictor associated with a significant coefficient when two predictors were added simultaneously.
* $p < .05$.

and content scales and how each set fares as a predictor of these criteria independent of the other. For each of the criteria listed in Tables 5 and 6, the first row of statistics indicates the performance of the clinical scale or scales included in the analyses independent of the content scales. Tables 7 and 8 provide equivalent information for the content scales.

Results for the two sets were very consistent. For the clinical scales, 19 out of 26 concurrent validity analyses were significant. The corresponding number for the content scales was 21 out of 26. Either set successfully predicted the majority of the criteria examined. Both sets were effective predictors of almost every chart review score. Both sets predicted admission depressive diagnoses and both admission and discharge psychotic diagnoses in both genders; they also predicted discharge depressive diagnoses in girls only. Neither set was very effective at predicting conduct disorder diagnoses. Finally, both sets were successful predictors of the HPRS component scores. For the content scales, the mean proportion of variance predicted, excluding diagnostic analyses, was .08 (*Mdn* = .08). For the clinical scales, the proportion of variance predicted was, on average, .07 (*Mdn* = .03).

Discussion

Overall, the statistics suggest two sets of scales that are approximately equal to each other in terms of their concurrent and

Table 6
Incremental Validity for the Content Scales in Girls

Criterion	Predictor	R^2	F or χ^2	*df*	Change analysis R^2	Change analysis F or χ^2	Change analysis *df*
HPRS components							
Negative affect	7	.01	3.8	1, 289			
	Anx	.02	2.2	2, 288	.00	0.59	1, 288
Depression	2	.07	20.3*	1, 289			
	Dep	.07	10.4*	2, 288	.00	0.31	1, 288
Diagnosis							
Admission							
Psychotic	8	.06	16.8*	1			
	Biz	.09	24.9*	2	.03	8.06*	1
Depression	2	.02	8.1*	1			
	Dep	.03	11.4*	2	.01	3.31	1
Conduct disorder	4, 9	.00	0.1	2			
	Con[a], Ang	.03	10.1*	4	.03	10.05*	2
Discharge							
Psychotic	8	.07	21.6*	1			
	Biz	.10	29.9*	2	.03	8.31*	1
Depression	2	.02	6.2*	1			
	Dep	.03	10.7*	2	.01	4.52*	1
Conduct disorder	4, 9	.00	0.2	2			
	Con, Ang	.01	3.6	4	.01	3.40	2
Chart clusters							
Self-harm	2	.03	11.3*	1, 321			
	Dep	.11	20.3*	2, 320	.08	28.50*	1, 320
Conduct	4, 9	.04	5.8*	2, 321			
	Con[a], Ang[a]	.13	11.8*	4, 319	.09	17.21*	2, 319
Bizarreness	8	.18	68.1*	1, 321			
	Biz	.19	38.1*	2, 320	.02	6.73*	1, 320
Obsessiveness	7	.03	11.0*	1, 322			
	Obs	.03	5.6*	2, 321	.00	0.33	1, 321
Depression	2	.14	53.0*	1, 323			
	Dep	.15	28.6*	2, 322	.01	3.79*	1, 322

Note. If two numbers are listed for *df*, the previous statistic is an *F* value; if one number is listed for *df*, the previous statistic is a chi-square value. R^2 refers to McFadden's rho squared. HPRS = Hopkins Psychiatric Rating Scale; 7 = Scale 7: Psychasthenia; Anx = Anxiety; 2 = Scale 2: Depression; Dep = Depression; 8 = Scale 8: Schizophrenia; Biz = Bizarre Mentation; 4 = Scale 4: Psychopathic Deviate; 9 = Scale 9: Hypomania; Con = Conduct Problems; Ang = Anger; Obs = Obsessiveness.
[a] Indicates a predictor associated with a significant coefficient when two predictors were added simultaneously.
* $p < .05$.

incremental validity. The number of significant incremental validity tests was slightly larger for the content scales than for the clinical scales (13 vs. 7), but the mean and median increments in validity were consistent. Results from the concurrent validity analyses also closely mirrored each other, with a slight advantage for the content scales over the clinical scales.

The largest difference appeared in the pattern of variables for which each set offered incremental validity. The content scales contributed more to the prediction of conduct problems when added to the clinical scales. Scale 7 tended to add more to the prediction of obsessiveness than did the Obsessiveness content scale. As expected on the basis of previous literature (e.g., Barthlow et al., 1999), there were some variations in outcomes across genders. However, the similarities between boys and girls were generally stronger than the differences. Specifically, in almost all instances in which an incremental validity test was significant among the boys, the same was true for the girls, though there were more significant outcomes among the girls than the boys.

One finding worth noting is the generally low level of incremental validity in this study. The mean proportion of variance predicted in the concurrent validity analyses was .08 for the content scales, corresponding to a mean multiple correlation of .28. The mean increment resulting from adding the clinical scales was .02, or a .04 increase in the mean corresponding correlation. This is consistent with evidence presented by Barthlow et al. (1999),

Table 7
Incremental Validity for the Clinical Scales in Boys

Criterion	Predictor	R^2	F or χ^2	*df*	Change analysis R^2	Change analysis F or χ^2	Change analysis *df*
HPRS components							
Negative affect	Anx	.02	5.3*	1, 285			
	7	.03	3.6*	2, 284	.01	2.04	1, 284
Depression	Dep	.03	8.7*	1, 285			
	2	.03	4.4*	2, 284	.01	3.51	1, 284
Diagnosis							
Admission							
Psychotic	Biz	.06	15.5*	1			
	8	.06	15.6*	2	.00	0.11	1
Depression	Dep	.01	5.1*	1			
	2	.02	6.0*	2	.01	0.89	1
Conduct disorder	Con[a], Ang	.01	4.6	2			
	4, 9	.01	4.8	4	.00	0.20	2
Discharge							
Psychotic	Biz	.03	7.3*	1			
	8	.03	7.4*	2	.00	0.09	1
Depression	Dep	.01	2.0	1			
	2	.01	2.1	2	.00	0.12	1
Conduct disorder	Con[a], Ang	.03	9.9*	2			
	4, 9	.03	11.5*	4	.00	1.67	2
Chart clusters							
Self-harm	Dep	.15	54.7*	1, 301			
	2	.17	30.3*	2, 300	.01	5.05*	1, 300
Conduct	Con, Ang	.07	11.8*	2, 299			
	4, 9	.07	5.9*	4, 297	.00	0.16	2, 297
Bizarreness	Biz	.17	60.4*	1, 302			
	8	.18	33.6*	2, 301	.02	5.89*	1, 301
Obsessiveness	Obs	.00	1.2	1, 301			
	7	.01	1.0	2, 300	.00	0.91	1, 300
Depression	Dep	.09	29.6*	1, 301			
	2	.14	24.0*	2, 300	.05	17.05*	1, 300

Note. If two numbers are listed for *df*, the previous statistic is an *F* value; if one number is listed for *df*, the previous statistic is a chi-square value. R^2 refers to McFadden's rho squared. HPRS = Hopkins Psychiatric Rating Scale; Anx = Anxiety; 7 = Scale 7: Psychasthenia; Dep = Depression; 2 = Scale 2: Depression; Biz = Bizarre Mentation; 8 = Scale 8: Schizophrenia; Con = Conduct Problems; Ang = Anger; 4 = Scale 4: Psychopathic Deviate; 9 = Scale 9: Hypomania; Obs = Obsessiveness.
[a] Indicates a predictor associated with a significant coefficient when two predictors were added simultaneously.
* $p < .05$.

who also examined relationships between MMPI–2 scales and clinician ratings.

Several factors can account for this finding. One is the sample's low rate of clinical elevations on the MMPI–A scales. Although this finding is typical in psychiatric populations (Archer et al., 2001), it is a pattern likely to dampen any effects. To evaluate this hypothesis, we repeated the analyses, this time eliminating cases from the sample if they had no clinical elevations on any of the 11 MMPI–A scales investigated in this study. Using this criterion rather than validity scale data to exclude cases had no consistent impact on the size of the effects. Consistency with recent findings for the MMPI–2 (Barthlow et al., 1999) further argues against lack of elevation as the predominant cause for this finding.

Another more likely possibility is that the corresponding clinical and content scales are relatively similar in terms of what they measure. In support of this hypothesis, it may be noted that the matched pairs of clinical and content scales tended to correlate fairly highly with each other: Correlations varied between .39 (for Scale 4 and the Anger content scale) and .78 (for Scale 8 and the Bizarre Mentation scale).

Given that the content scales offered relatively small increments in fit when used to predict clinically important variables, one may wonder whether these findings justify the use of the content scales

Table 8
Incremental Validity for the Clinical Scales in Girls

Criterion	Predictor	R^2	F or χ^2	*df*	Change analysis R^2	Change analysis F or χ^2	Change analysis *df*
HPRS components							
Negative affect	Anx	.01	4.1*	1, 289			
	7	.02	2.2	2, 288	.00	0.29	1, 288
Depression	Dep	.04	12.5*	1, 289			
	2	.07	10.4*	2, 288	.03	8.03*	1, 288
Diagnosis							
Admission							
Psychotic	Biz	.09	24.7*	1			
	8	.09	24.9*	2	.00	0.18	1
Depression	Dep	.03	10.7*	1			
	2	.03	11.4*	2	.00	0.71	1
Conduct disorder	Con[a], Ang	.02	5.9	2			
	4, 9	.03	10.1*	4	.01	4.16	2
Discharge							
Psychotic	Biz	.09	29.3*	1			
	8	.10	29.9*	2	.01	0.57	1
Depression	Dep	.03	10.6*	1			
	2	.03	10.7*	2	.00	0.15	1
Conduct disorder	Con, Ang	.01	2.0	2			
	4, 9	.01	3.6	4	.00	1.66	2
Chart clusters							
Self-harm	Dep	.11	39.6*	1, 321			
	2	.11	20.3*	2, 320	.00	1.08	1, 320
Conduct	Con[a], Ang[a]	.13	23.4*	2, 321			
	4, 9	.13	11.8*	4, 319	.00	0.37	2, 319
Bizarreness	Biz	.17	63.4*	1, 321			
	8	.19	38.1*	2, 320	.03	10.69*	1, 320
Obsessiveness	Obs	.01	4.4*	1, 322			
	7	.03	5.6*	2, 321	.02	6.65*	1, 321
Depression	Dep	.11	38.7*	1, 323			
	2	.15	28.6*	2, 322	.04	16.69*	1, 322

Note. If two numbers are listed for *df*, the previous statistic is an *F* value; if one number is listed for *df*, the previous statistic is a chi-square value. R^2 refers to McFadden's rho squared. HPRS = Hopkins Psychiatric Rating Scale; Anx = Anxiety; 7 = Scale 7: Psychasthenia; Dep = Depression; 2 = Scale 2: Depression; Biz = Bizarre Mentation; 8 = Scale 8: Schizophrenia; Con = Conduct Problems; Ang = Anger; 4 = Scale 4: Psychopathic Deviate; 9 = Scale 9: Hypomania; Obs = Obsessiveness.
[a] Indicates a predictor associated with a significant coefficient when two predictors were added simultaneously.
* $p < .05$.

for incremental clinical purposes. The answer depends on whether the question arises during the administration or scoring of the inventory. If the adolescent's willingness or capacity to complete the entire inventory is questionable, the clinician may choose to think twice about requiring the completion of all the items as is needed to score the content scales in addition to the clinical scales. Once the inventory is completed, though, the answer depends on whether it is to be scored by computer or by hand. If the former, the content scale scores are generated at no additional cost to the clinician, and so even small increments in validity are cost-effective.

In general, the results demonstrate the content scales are roughly equivalent, if not slightly superior, to the clinical scales in terms of both concurrent validity and incremental validity. Among the content scales, results were particularly supportive for using the Bizarre Mentation and Depression scales, and to a lesser extent the Conduct Problems and Anger scales, as either adjuncts or alternatives to the traditional clinical scales. Results were similarly supportive of clinical scales 2, 7, and 8. Though the overall results suggest rough equivalence among the two sets, it must be remembered that the analyses were limited to a subset of the clinical and content scales. It is not really accurate to portray these results as indicative of these scales as a whole, despite the fact that we have done so in this article for the sake of brevity. These results merit cross-validation with additional predictors, with other clinically important criteria, and with other populations besides inpatients.

References

Archer, R. P., Aiduk, R., Griffin, R., & Elkins, D. E. (1996). Incremental validity of the MMPI–2 content scales in a psychiatric sample. *Assessment, 3,* 79–90.

Archer, R. P., Handel, R. W., & Lynch, K. D. (2001). The effectiveness of MMPI–A items in discriminating between normative and clinical samples. *Journal of Personality Assessment, 77,* 420–435.

Archer, R. P., & Krishnamurthy, R. (1997). MMPI–A and Rorschach indices related to depression and conduct disorder: An evaluation of the incremental validity hypothesis. *Journal of Personality Assessment, 69,* 517–533.

Arita, A. A., & Baer, R. A. (1998). Validity of selected MMPI–A content scales. *Psychological Assessment, 10,* 59–63.

Barthlow, D. L., Graham, J. R., Ben-Porath, Y. S., & McNulty, J. L. (1999). Incremental validity of the MMPI–2 content scales in an outpatient mental health setting. *Psychological Assessment, 11,* 39–47.

Ben-Porath, Y. S., Butcher, J. N., & Graham, J. R. (1991). Contribution of the MMPI–2 content scales to the differential diagnosis of schizophrenia and major depression. *Psychological Assessment, 3,* 634–640.

Ben-Porath, Y. S., McCully, E., & Almagor, M. (1993). Incremental validity of the MMPI–2 content scales in the assessment of personality and psychopathology by self-report. *Journal of Personality Assessment, 61,* 557–575.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *MMPI–2: Manual for administering and scoring.* Minneapolis: University of Minnesota Press.

Butcher, J. N., Graham, J. R., Williams, C. L., & Ben-Porath, Y. S. (1990). *Development and use of the MMPI–2 content scales.* Minneapolis: University of Minnesota Press.

Butcher, J. N., Williams, C. L., Graham, J. R., Archer, R. P., Tellegen, A., Ben-Porath, Y. S., & Kaemmer, B. (1992). *MMPI–A manual for administration, scoring, and interpretation.* Minneapolis: University of Minnesota Press.

Cashel, M. L., Rogers, R., Sewell, K. W., & Holliman, N. B. (1998). Preliminary validation of the MMPI–A for a male delinquent sample: An investigation of clinical correlates and discriminant validity. *Journal of Personality Assessment, 71,* 49–69.

Cull, J. G., & Gill, W. S. (1988). *Suicide Probability Scale* (*SPS*) *manual.* Los Angeles: Western Psychological Services.

Derogatis, L. R. (1983). *SCL–90–R administration, scoring, and procedures manual.* Towson, MD: Clinical Psychometric Research.

Hathaway, S. R., & McKinley, J. C. (1983). *Manual for the administration and scoring of the MMPI.* Minneapolis, MN: National Computer Systems.

Hollrah, J. L., Schlottmann, R. S., Scott, A. B., & Brunetti, D. G. (1995). Validity of the MMPI subtle items. *Journal of Personality Assessment, 65,* 278–299.

Jackson, D. N. (1971). The dynamics of structured personality tests: 1971. *Psychological Review, 78,* 229–248.

Kopper, B. A., Osman, A., Osman, J. R., & Hoffman, J. (1998). Clinical utility of the MMPI–A content scales and Harris–Lingoes subscales in the assessment of suicidal risk factors in psychiatric adolescents. *Journal of Clinical Psychology, 54,* 191–200.

Overall, J. E., & Gorham, D. R. (1962). The Brief Psychiatric Rating Scale. *Psychological Reports, 10,* 799–812.

Pena, L. M., Megargee, E. I., & Brody, E. (1996). MMPI–A patterns of male juvenile delinquents. *Psychological Assessment, 8,* 388–397.

Tabachnik, B. G., & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.). New York: HarperCollins.

Weed, N. C., Ben-Porath, Y. S., & Butcher, J. N. (1990). Failure of Wiener and Harmon Minnesota Multiphasic Personality Inventory (MMPI) subtle scales as personality descriptors and as validity indicators. *Psychological Assessment, 2,* 281–285.

Wiggins, J. S. (1966). Substantive dimensions of self-report in the MMPI item pool. *Psychological Monographs, 80*(22, Whole No. 630).

Williams, C. L., & Butcher J. N. (1989). An MMPI study of adolescents: I. Empirical validity of the standard scales. *Psychological Assessment, 1,* 251–259.

Williams, C. L., Butcher, J. N., Ben-Porath, Y. S., & Graham, J. R. (1992). *MMPI–A content scales: Assessing psychopathology in adolescents.* Minneapolis: University of Minnesota Press.

Received April 10, 2001
Revision received May 20, 2002
Accepted June 12, 2002 ■