This excerpt is taken from:

Manual for the ASEBA Adult Forms & Profiles

For Ages 18-59

- ■ **Adult Self-Report**
- ■ **Adult Behavior Checklist**

An Integrated System of Multi-informant Assessment

Thomas M. Achenbach, University of Vermont & Leslie A. Rescorla, Bryn Mawr College

Chapter 9

Reliability, Internal Consistency, Cross-Informant Agreement, and Stability

Reliability refers to agreement between repeated assessments of characteristics when the characteristics themselves are expected to remain constant. When instruments such as the ASR and ABCL are completed, it is important to know the degree to which scale scores remain consistent over periods when the adaptive characteristics and problems of the people who are assessed are not likely to change much. In this chapter, we present test-retest reliabilities for the ASR and ABCL scale scores over 1-week intervals.

Another property of scale scores is their *internal consistency.* This refers to the degree to which the items of a scale are correlated with each other. Internal consistency is sometimes called *split-half reliability*, because it can be estimated by correlating the sum of scores on half the items of a scale with the sum of scores on the other half. However, internal consistency among the items of a scale scored on a single occasion cannot tell us the reliability with which the scale will produce the same results on different occasions.

Although test-retest reliability and internal consistency are typically viewed as psychometric properties of the scales themselves, reports of adaptive functioning and problems inevitably depend on the informants' perspectives. Assessment of adults has traditionally relied primarily on the adults' self-reports via interviews, tests, and questionnaires. Although it has long been clear that multi-informant reports are needed for assessing children and adolescents, the need for multi-informant assessment of adults is becoming evident, as well (Klonsky et al. 2002; Meyer, 2002; Meyer et al., 2001). As illustrated in preceding chapters, the ASR and ABCL are designed to obtain and systematically compare data from self-reports and reports by other people for the assessment of individuals and for research. In this chapter, we will present findings for *cross-informant agreement* between scale scores obtained from self-reports and reports by others.

A further aspect of scale scores is their *stability* when the same informants complete forms over intervals long enough for behavior to change. Stability can be affected by many factors, including aging, therapeutic interventions, and important life events such as marriage, job changes, and traumas. Nevertheless, data on the stability of scale scores for substantial samples of adults can provide reference points for the typical level of those scales' stability.

TEST-RETEST RELIABILITIES OF SCALE SCORES

To assess reliability in both the rank ordering and magnitude of scale scores, we computed test-retest Pearson correlations (*r)* and *t* tests of differences between ASR ratings and between ABCL ratings on two occasions. The test-retest reliability samples were obtained by requesting participants in our 1999 National Survey to complete their respective forms twice at intervals averaging 7 days. Because the ASR for ages 30-59 used in the 1999 National Survey did not include the Education items, we report the test-retest *r* that was reported for the Education scale in the Manual for the YASR and YABCL (Achenbach, 1997). The items and scoring were the same as for the current Education scale. As detailed in the Manual for the YASR and YABCL, the data were obtained from 232 young adults who completed the YASR twice at a mean of 7 days. They included 50 participants in our previous national sample, 11 participants in a follow-up of adolescent mental health services, 123 American students, and 48 Turkish students who completed a Turkish translation of the ASR.

Test-Retest Correlations

As shown in Table 9-1, reliability was generally very high, with all test-retest *r*s being significant at *p*<.01 and most being in the .80's and .90's. The mean

Table 9-1
One-Week Test-Retest Reliabilities, Alphas, and Cross-Informant Correlations[a]

Scales	Test-Retest r		Alpha Coefficients		Cross-Informant r
	ASR	ABCL	ASR	ABCL	ASR x ABCL
Adaptive Functioning	N = 51	54	295	402	1,196
Friends	.82	.82	.69	.67	.48
Spouse/Partner	.85	.84	.78	.76	.54
Family[b]	.74	NA	f	NA	NA
Job[b]	.71	NA	.60	NA	NA
Education[b]	.80	NA	.51	NA	NA
Mean Adaptive[b]	.79	NA	f	NA	NA
Substance Use					
Tobacco	.93	.94	f	f	.79
Alcohol	.82	.92	f	f	.67
Drugs	.99	.87	f	f	.42
Mean Substance Use	.96	.91	f	f	.69
Critical Items	.87[d]	.84	.74	.81	.43
Empirically Based					
Anxious/Depressed	.87[d]	.75	.88	.90	.44
Withdrawn	.87[d, e]	.73	.78	.81	.31
Somatic Complaints	.78[d, e]	.88[d, e]	.82	.77	.36
Thought Problems	.91	.86	.51	.71	.30
Attention Problems	.91[d]	.87	.87	.88	.38
Aggressive Behavior	.87[d]	.87	.83	.91	.42
Rule-Breaking Behavior	.79[d]	.86	.86	.86	.46
Intrusive	.83[d, e]	.88	.72	.70	.38
Internalizing	.89[d]	.80	.93	.92	.43
Externalizing	.91[d]	.92	.89	.93	.44
Total Problems	.94[d]	.92	.97	.97	.42
Mean r[c] and mean alpha	.88[c]	.86[c]	.83	.85	.40[c]
DSM-Oriented					
Depressive Problems	.86[d, e]	.83[d, e]	.82	.88	.43
Anxiety Problems	.86[d]	.77	.68	.70	.41
Somatic Problems	.77	.87[d, e]	.79	.75	.34
Avoidant Personality Problems	.85[d]	.75	.74	.77	.35
AD/H Problems	.84[d]	.89	.84	.83	.35
Inattention	.80[d]	.81	.79	.75	.36
Hyperactivity-Impulsivity	.81	.91	.76	.76	.34
Antisocial Personality Problems	.84[d]	.87	.79	.87	.42
Mean r[c] and mean alpha	.83[c]	.85[c]	.78	.79	.38[c]
Mean Q correlation between problem items	NA	NA	NA	NA	.30

[a]Test-retest samples were 1999 National Survey respondents interviewed at mean intervals of 7 days, except Education scale for which data were obtained from 232 participants in previous national survey and other sources (Achenbach, 1997) described in the text. Cronbach's alpha was computed for referred and nonreferred samples of ages 30-59 described in Chapter 10, except Education which was computed for age 18-29 YASRs (*N*=305). Cross-informant *r*s were for National Survey respondents. All *r*s were *p*<.01. *N*s vary because adaptive scales were not relevant for all participants. [b]Scales that are only on ASR. [c]Mean *r* computed by *z* transformation. [d]Time 1 > Time 2, at *p*<.05, by *t* tests. [e]When corrected for the number of comparisons, Time 1 vs. Time 2 difference was not significant (Sakoda et al., 1954). [f]Alpha was not applicable because these represent single items or a mean score.

r for the empirically based problem scales was .88 on the ASR and .86 on the ABCL, while the *r*s for Total Problems were .94 and .92, respectively. For the DSM-oriented scales, the mean *r* was .83 on the ASR and .85 on the ABCL.

There were significant (*p*<.05) declines in scores on the scales that are marked with superscript *d* in Table 9-1. Four of the significant changes in each column would be expected by chance, based on the number of analyses that were done using a *p*<.05 protection level (Sakoda, Cohen, & Beall, 1954). Superscript *e* indicates the differences that were most likely to be significant by chance, because they yielded the smallest *t* values.

Called the *test-retest attenuation effect*, the tendency for people to report fewer problems at a second assessment is often found in rating forms, questionnaires, and structured psychiatric interviews for children, adolescents, and adults (e.g., Achenbach & Rescorla, 2001; Helzer, Spitznagel, & McEvoy, 1987; Roberts, Solovitz, Chen, & Casat, 1996; Vandiver & Sher, 1991). Note that this effect is not due to regression toward the mean, because it is found in samples whose initial scores are not especially high. As Table 9-1 shows, 17 of the 32 test-retest comparisons for the ASR showed significant changes in mean scores. The four marked with superscript *e* would be expected by chance. The 3 significant declines in ABCL scores were fewer than the 4 expected by chance in 27 analyses (Sakoda et al., 1954). None of the adaptive functioning or substance use scales showed significant changes in mean scores over the 1-week interval.

Test-Retest Reliability for Turkish vs. English Administrations

To provide an especially stringent assessment of test-retest reliability, Erol (unpublished data) had 50 bilingual Turks complete Turkish and English versions of the YASR in counterbalanced order at intervals averaging one week. The mean test-retest *r* for the 11 YASR empirically based problem scales was .88, for adaptive functioning was .79, and for substance use was .81. These *r*s did not differ significantly from the corresponding *r*s shown in Table 9-1 for Americans who completed ASEBA forms only in their native language. English vs. Turkish x Time 1 vs. Time 2 analyses of variance (ANOVAs) showed no significant differences between scores on any scales of the English vs. Turkish YASRs (all *p* >.05 after Bonferroni correction). The Turkish subjects' responses were thus not affected by the language in which they read the YASR.

INTERNAL CONSISTENCY OF SCALE SCORES

High internal consistency is often considered to be desirable and is sometimes interpreted as indicating that a scale is highly reliable. However, because internal consistency is computed on the basis of item scores that were all obtained at the same time, it cannot really tell us the test-retest reliability of a scale. Furthermore, some scales with very high internal consistency may not be as valid as some scales with lower internal consistency.

As an example, if a scale consists of 20 versions of the same question, it should have very high internal consistency, because respondents should give similar answers to the 20 versions of the question. However, such a scale would usually be less valid than a scale that used 20 different questions to assess the same phenomenon. Because each of the 20 different questions is likely to tap different aspects of the target phenomenon and to be subject to different errors of measurement, the 20 different questions are likely to provide better measurement despite lower internal consistency than a scale that used 20 versions of a single question.

As detailed in Chapter 7, our syndrome scales were derived from factor analyses of the correlations among ASEBA items. The composition of the scales is therefore based on internal consistencies among certain subsets of items. Nevertheless, because some users may wish to know the degree of internal consistency of our scales, Table 9-1 displays Cronbach's (1951) *alpha* for each scale. Alpha represents the mean of the correlations between all sets of half the items comprising a scale. Alpha tends to be directly related to the length of the scale, because half the items of a short scale provide a less stable measure than half the items of a long scale.

As Table 9-1 shows, the alphas for the adaptive functioning scales were moderately high, ranging from

.60 to .78, except the Education scale, whose alpha was .51. These alphas are reasonable for scales that have relatively few items.

For the empirically based problem scales, the alphas ranged from .51 to .97. The only alpha < .70 was on the ASR Thought Problems syndrome, which is comprised of low-prevalence items. For the DSM-oriented scales, the alphas ranged from .68 to .88. The only alpha < .70 was on the ASR Anxiety Problems scale.

CROSS-INFORMANT AGREEMENT

Table 9-1 displays Pearson rs between raw scale scores for ASRs and ABCLs completed by participants in our National Survey. The ASRs were completed by the participants themselves, whereas the ABCLs were completed by people nominated by the participants. Those completing the ABCLs included spouses, partners, parents, relatives, children, friends, and roommates of the participants.

The cross-informant correlations ranged from .30 for the Thought Problems syndrome to .79 for the Tobacco scale, all $p < .001$. The mean cross-informant r was .40 for the empirically based problem scales and .38 for the DSM-oriented scales. The median r was .42 for the empirically based problem scales, while it was .36 for the DSM-oriented scales. This was the same as the median r of .36 found by Klonsky et al. (2002) for 10 studies of rs between self-reports and other-reports of DSM personality disorders.

In addition to Pearson rs between ASR and ABCL scale scores, the bottom line of Table 9-1 displays the mean of the Q correlations between the 0-1-2 scores obtained on the problem items of ASRs completed by participants in our National Survey and on the counterpart items of ABCLs completed for the sample participants by people who knew them. Each Q correlation reflects the degree of agreement between the pattern of 0-1-2 scores for ASR problem items and ABCL problem items scored for the same individual. Like Pearson r, Q correlations can range from −1.00, which indicates total disagreement, to +1.00, which indicates perfect agreement between two sets of scores for the

same individual. The mean Q correlation of .30 indicates a modest level of agreement. If you use the Ages 18-59 Module to score an ASR and up to seven ABCLs for the same individual, you can have the module display Q correlations between the ASR and each of the ABCLs. If you choose this option, the module will also display the mean Q correlation shown in Table 9-1, plus the 25th percentile Q correlation and the 75th percentile Q correlation. The 25th percentile Q correlation = .17, while the 75th percentile Q correlation = .43. If a particular Q correlation is < .17, the printout will state that it is below average. If the Q correlation is > .43, the printout will state that it is above average.

STABILITY OF SCALE SCORES

To assess long-term stability in the rank ordering of scale scores, we computed rs between the 1997 scales scored from YASRs completed by 484 young adults and scales scored from YABCLs completed by 587 parents at intervals averaging 39 months for YASRs and 44 months for YABCLs. As reported in Chapter 12, the 1997 YASR and YABCL scores correlated highly with the current ASR and ABCL scales. The YASR respondents included 366 from the national sample that was used to norm the young adult instruments (Achenbach, 1997), plus 118 from a follow-up study of former clients of the child/adolescent mental health services at the University of Vermont. All YABCL respondents were parents of participants in the national sample, most of whom completed the forms via home interview in 1992 and by mail in 1995-96. At each assessment, a few respondents provided data via telephone interviews.

Table 9-2 shows that the long-term rs averaged .58 for the YASR and .60 for the YABCL. As Table 9-2 shows, there were significant declines in scores on five YASR and YABCL scales. However, the largest decline on a YASR scale (Attention Problems) accounted for only 0.37% of variance, while the largest decline on a YABCL scale (Intrusive) accounted for only 0.41% of variance, both of which fell below Cohen's (1988) criteria of 1 to 5.9% of variance for small ES.

Over a 2-year interval, Ferdinand and Verhulst (1995) obtained a mean stability r of .69 for YASR Total Problems scores among 528 18- to 23-year-

Table 9-2
Long-Term Stabilities of Empirically Based Problem Scales

Scales	*YASR*	*YABCL*	*Mean of YASR & YABCL r[b]*
	N = 484	N = 587	
Anxious/Depressed	.63	.66[a]	.65
Withdrawn	.54	.57[a]	.56
Somatic Complaints	.54	.36	.46
Thought Problems	.42	.30	.36
Attention Problems	.54[a]	.67[a]	.61
Aggressive Behavior	.61[a]	.68	.65
Rule-Breaking Behavior	.56	.57	.57
Intrusive	.57[a]	.61[a]	.59
Internalizing	.62	.67[a]	.65
Externalizing	.63[a]	.70	.67
Total Problems	.65[a]	.68	.67
Mean r[b]	.58	.60	.59

Note. Mean intervals were 39 months for YASRs and 44 months for YABCLs. All Pearson *r*s were significant at *p* <.001.
[a]Time 1 > Time 2, *p* <.05 by *t* test, after Bonferroni correction.
[b]Computed by *z* transformation.

olds in a Dutch general population sample. A multivariate analysis of variance (MANOVA) that included all problem scales showed no significant change in scores over the 2-year period. Over intervals averaging 8 months, YASR stabilities measured by Spearman correlations for 131 young adults in a Dutch general population sample were .79 for Total Problems, .79 for Internalizing, and .77 for Externalizing (Ferdinand, van der Reijden, Verhulst, Nienhuis, and Giel, 1995).

In the same Dutch general population sample, Hofstra van der Ende, and Verhulst (2001) computed *r*s between problem scores on YSRs completed by the participants at ages 15 to 19 years and YASRs completed 10 years later at ages 25 to 29 years. For Total Problems scores, the 10-year *r*s were .53 for males (*N* = 177) and .43 for females (*N* = 235). Although these *r*s were limited by the fact that most of the participants completed the YSR at younger ages than the YASR was designed for and that the YSR and YASR Total Problems scores were not based on exactly the same items, they indicate substantial stability in self-reported problems even over 10 years.

SUMMARY

The test-retest reliability of ASEBA adult forms was supported by 1-week test-retest *r*s that were in the .80's and .90's for most scales, with none < .71.

Good internal consistency was found for most scales, with mean alpha coefficients on the ASR and ABCL of .83 and .85 for the empirically based problem scales and .78 and .79 for the DSM-oriented scales.

Cross-informant *r*s between ASR and ABCL scores averaged .69 for substance use, .40 for the empirically based problem scales, and .38 for the DSM-oriented scales. The mean *Q* correlation between ASR and ABCL problem items was .30. Consistent with findings from other instruments, the modest size of the cross-informant correlations indicates the need to obtain data from multiple informants whenever possible.

Substantial long-term stability of scores was indicated by *r*s ranging from .69 over a 2-year interval, to .58-.60 over 39-44 months, and .43-.53 over 10 years.

Chapter 10

Validity

Validity refers to the accuracy with which instruments assess what they are supposed to assess. ASEBA instruments serve many purposes, and their validity can be evaluated in multiple ways. A fundamental purpose of the ASR and ABCL is to aid in identifying needs for help with behavioral, emotional, and social problems and adaptive functioning. The ASR and ABCL also provide well-differentiated pictures of people's functioning in terms of specific problems and adaptive characteristics, aggregations of related problems into empirically based and DSM-oriented scales, and broader aggregations of items that encompass diverse aspects of functioning. In this chapter, we present evidence for the *content validity, criterion-related validity,* and *construct validity* of the ASR and ABCL.

CONTENT VALIDITY

The most basic kind of validity is *content validity*, which is the degree to which an instrument's content includes what the instrument is intended to assess.

Problem Items

The ASR and ABCL problem items are products of a long process of development, testing, and refinement on the basis of research and practical experience. The process began with the selection of items for assessing children and youth on the basis of extensive literature searches, consultation with relevant professionals, and repeated pilot testing in a variety of samples (Achenbach, 1965, 1966; Achenbach & Edelbrock, 1983; Achenbach & Lewis, 1971). Applications of the ASEBA approach to assessment of adults began in the 1980s with development of the YASR and YABCL to assess adults who had previously been assessed with ASEBA school-age instruments (Achenbach, 1991a; Achenbach, Howell, McConaughy, & Stanger, 1995c; Stanger, MacDonald, McConaughy, & Achenbach, 1996).

The *Manual for the Young Adult Self-Report and Young Adult Behavior Checklist* (Achenbach, 1997) provides details of the refinement and testing of the items and scales for the young adult forms and profiles. The development of the forms for ages 30-59 began in the 1990s with the addition of items to the young adult forms, further testing and revision, and use of the forms for ages 30-59 in our 1999 National Survey and in various clinical and nonclinical settings. As described in Chapter 7 of this *Manual*, our separate factor analyses of the forms for ages 18-29 and for ages 30-59 yielded such similar results that the data from both sets of forms were combined in the final analyses from which the ASR and ABCL syndromes were derived.

As documented in Chapter 12, the current ASR and ABCL syndromes are quite similar to those derived from the YASR and YABCL (Achenbach, 1997) via different factor analytic methodology, and scores on the current syndromes correlate highly with scores on the 1997 versions. All problem items of the current versions of the adult forms were *(a)* scored significantly higher on one or both forms for adults referred for mental health or substance use services than for demographically similar adults who had not been referred for such services in the preceding 12 months (details in Chapter 11), and/or *(b)* loaded significantly on empirically based syndromes (details in Chapter 7), and/or *(c)* were identified by the expert panel as being very consistent with DSM-IV diagnostic categories (details in Chapter 4).

Adaptive Functioning Items

The adaptive functioning items were hypothesized to reflect aspects of functioning that are important for successful adaptation in various areas. Most of the items were found to discriminate significantly between referred and nonreferred young adults (Achenbach,

1997). A few items have been made more differentiated and a few items have been added since the 1997 version of the YASR, as detailed in Chapter 12. The ASR adaptive functioning items pertaining to friends and family are relevant to nearly all adults, although the family items are not included on the ABCL, because many informants may not be knowledgeable about them. The items pertaining to spouse/partner, job, and education are completed only for adults for whom they have been relevant in the preceding 6 months.

Although included in the ASR Job section, item *IV.E. I am satisfied with my job situation,* is not scored on the Job scale. It is not scored because it was unexpectedly found to be rated significantly higher by referred than nonreferred respondents in the samples analyzed for this *Manual*, whereas it was rated significantly higher by nonreferred than referred young adults in our previous analyses of the YASR (Achenbach, 1997). Because of these contradictory findings, we decided to omit item *IV.E* from the Job scale, although it may be clinically useful. As detailed in Chapter 11, almost all other adaptive functioning items were rated significantly more favorably for nonreferred than referred adults on one or both adult forms.

In summary, the content validity of the ASR and ABCL items has been supported by a long process of item development, testing, and revision, as well as by findings that most of the items retained for scoring on scales discriminate significantly between demographically similar referred and nonreferred adults.

CRITERION-RELATED VALIDITY OF SCALE SCORES

Criterion-related validity refers to the strength of association between a particular measure, such as a scale scored from an ASEBA form, and an external criterion for characteristics that the scale is intended to assess. In the preceding section, we mentioned that most of the items retained for ASR and ABCL scales discriminated significantly between referred and nonreferred adults on one or both forms. Here we focus on associations between scales comprising particular sets of ASEBA items and external criterion vari-

ables. We will first present new validity evidence based on analyses done for this *Manual.* We will then summarize validity evidence from other sources.

Demographically Similar Samples of Referred and Nonreferred Adults

To test the ability of each ASEBA scale to discriminate between referred and nonreferred adults, we constructed referred and nonreferred samples that were matched for gender and were similar in age distributions. The referred adults came from 17 mental health and substance abuse treatment settings. In addition, to augment the sample of ABCLs, we used ABCLs that were completed for participants in our 1999 National Survey who reported receiving mental health or substance abuse services in the preceding 12 months. The ASRs and ABCLs for nonreferred adults were for participants in our 1999 National Survey who reported that they had not received mental health or substance abuse services in the preceding 12 months. We controlled for differences in age and ethnicity by treating them as covariates in ANCOVAs and as independent variables in multiple regression analyses. Table 10-1 summarizes characteristics of the referred and nonreferred samples.

Multiple Regression Analyses of Adaptive Functioning Scales

To test the associations of referral status and demographic variables with scale scores, we used a structural equation modeling (SEM) approach whereby we regressed the raw scores for a scale (the dependent variable) on the independent variables of referral status, gender, age (dichotomized as 18-35 vs. 36-59), and nonLatino white vs. other ethnicity. We entered all independent variables simultaneously to test the predictive power of each independent variable with the others partialed out.

For each adaptive scale, Table 10-2 displays the percentage of variance uniquely accounted for by referral status, with the effects of age, gender, and ethnicity partialed out. Cohen's (1988) criteria for effect sizes (ES) in multiple regression are as follows: Small = 2-13%; medium = 13-26%; and large >26%. The ES for referral status were large for 5 of the 6

Table 10-1
Characteristics of Demographically Similar Referred vs. Nonreferred Samples

	ASR		*ABCL*	
Characteristics	*Ref.*	*Nonref.*	*Ref.*	*Nonref.*
	N = 846	*N* = 846	*N* = 447	*N* = 447
Gender				
Men	45%	45%	53%	53%
Women	55%	55%	47%	47%
Age				
Mean	26.1	29.9	30.6	31.9
SD	8.9	9.5	12.8	11.9
Education[a]				
Mean	4.2	4.0	3.7	3.9
SD	1.8	1.9	1.9	1.9
SES[b]				
Mean	1.7	2.0	1.7	1.9
SD	0.7	0.7	0.8	0.7
Ethnicity				
NonLatino White	92%	66%	84%	69%
African American	3%	17%	7%	16%
Latino	2%	11%	5%	11%
Mixed or Other	4%	7%	4%	5%

[a]Education was scored: 1 = No high school diploma or General Equivalency Diploma (GED); 2 = GED; 3 = High school graduate; 4 = Some college; 5 = Associate's degree; 6 = Bachelor's or RN degree; 7 = Some graduate school; 8 = Master's degree; 9 = Doctoral or Law degree.
[b]SES was scored 1 = lower, 2 = middle, 3 = upper, based on an updated version of Hollingshead's (1975) 9-step scale for the occupation of the spouse/partner holding the higher status job: Hollingshead scores 1.0-3.9 = lower; 4.0-6.9 = middle; 7.0-9.0 = upper; we assigned 2-digit codes because occupations that were not clearly scorable were given the mean of their most likely scores.

ASR scales and were small for the ASR Family scale and for both ABCL adaptive functioning scales. Thus, after partialing out demographic variance, referral status accounted for significant variance in all the adaptive scales of the ASR. Figure 10-1 graphically displays the mean scores on each adaptive scale.

Demographic Effects. Scores were significantly higher for ages 36-59 than 18-35 on the ASR Job and Mean Adaptive scales, but, as Table 10-2 shows, the effects were small according to Cohen's (1988)

criteria. On the ABCL Friends scale, women obtained significantly higher scores than men, while on the ASR Friends scale, nonwhites scored themselves significantly higher than whites. However, these differences did not exceed chance expectations for the number of analyses (Sakoda et al., 1954).

Multiple Regression Analyses of Problem Scales

As we did for the adaptive functioning scales, we used SEM to regress the raw scores of each problem

Table 10-2
Percent of Variance Accounted for by Significant (*p* <.01) Effects of Referral Status and Age on Adaptive Functioning Scores in Multiple Regressions

Scales	Ref Status[a]		Age[b]	
	ASR	ABCL	ASR	ABCL
	N = 1,696	296	1,696	296
Friends	14	2	—	—
Spouse/Partner	16	3	—	—
Family	4	NA	—	NA
Job	37	NA	4[O]	NA
Education	27	NA	NA	NA
Mean Adaptive	50	NA	3[O,c]	NA

*Note. N*s vary because some scales were not relevant to some participants. Analyses were multiple linear regressions of raw scale scores on referral status, age, gender, and nonLatino white vs. other ethnicity. Ethnicity and gender each had one significant effect, which did not exceed chance expectations according to Sakoda et al. (1954).

[a]All scale scores were significantly (*p* = .000) higher for nonreferred than referred adults.

[b]O = ages 36-59 scored higher than ages 18-35.

[c]Not significant when corrected for number of analyses (Sakoda et al., 1954). Because all differences between nonreferred and referred adults were *p* = .000, none were apt to be chance differences.

scale on the independent variables of referral status, age (18-35 vs. 36-59), gender, and nonLatino white vs. other ethnicity. These analyses were done for the 3 substance use scales, mean substance use, Critical Items, 8 syndromes, Internalizing, Externalizing, Total Problems, 6 DSM-oriented scales, and 2 AD/H subscales.

Referral Status Effects. As Table 10-3 shows, the effects of referral status greatly outweighed the effects of demographic differences on most scales. Referred adults obtained significantly higher (*p*<.01) scores on all problem item scales except the ABCL Intrusive scale. The effects of referral status were larger on most of the ASR scales than on the counterpart ABCL scales. The largest effects of referral status were on the ASR DSM-oriented Depressive Problems scale (22% ES), the ASR Critical Items scale (20% ES), and the ASR Anxious/Depressed syndrome (17% ES). Figure 10-2 graphically displays the mean scores on each problem scale.

Demographic Effects. There were no significant effects of ethnicity on any problem scale. As Table 10-3 shows, there were numerous small effects of age, mostly reflecting higher scores for ages 18-35 than for ages 36-59. Gender effects were more numerous on the ASR than the ABCL, with each gender obtaining higher scores on some scales.

CLASSIFICATION OF PEOPLE ACCORDING TO CLINICAL CUTPOINTS

The regressions reported in the preceding section showed that nearly all quantitative scale scores discriminated significantly (*p*<.01) between referred and nonreferred people. Beside the quantitative scores, each scale has cutpoints for distinguishing categorically between the normal and clinical range. The cutpoints for the different scales were presented in Chapters 6, 7, and 8.

For some clinical and research purposes, users may wish to distinguish between people who are in the normal vs. clinical range according to the cutpoints. Because categorical distinctions are usually least reliable for individuals who score close to the border of a cat-
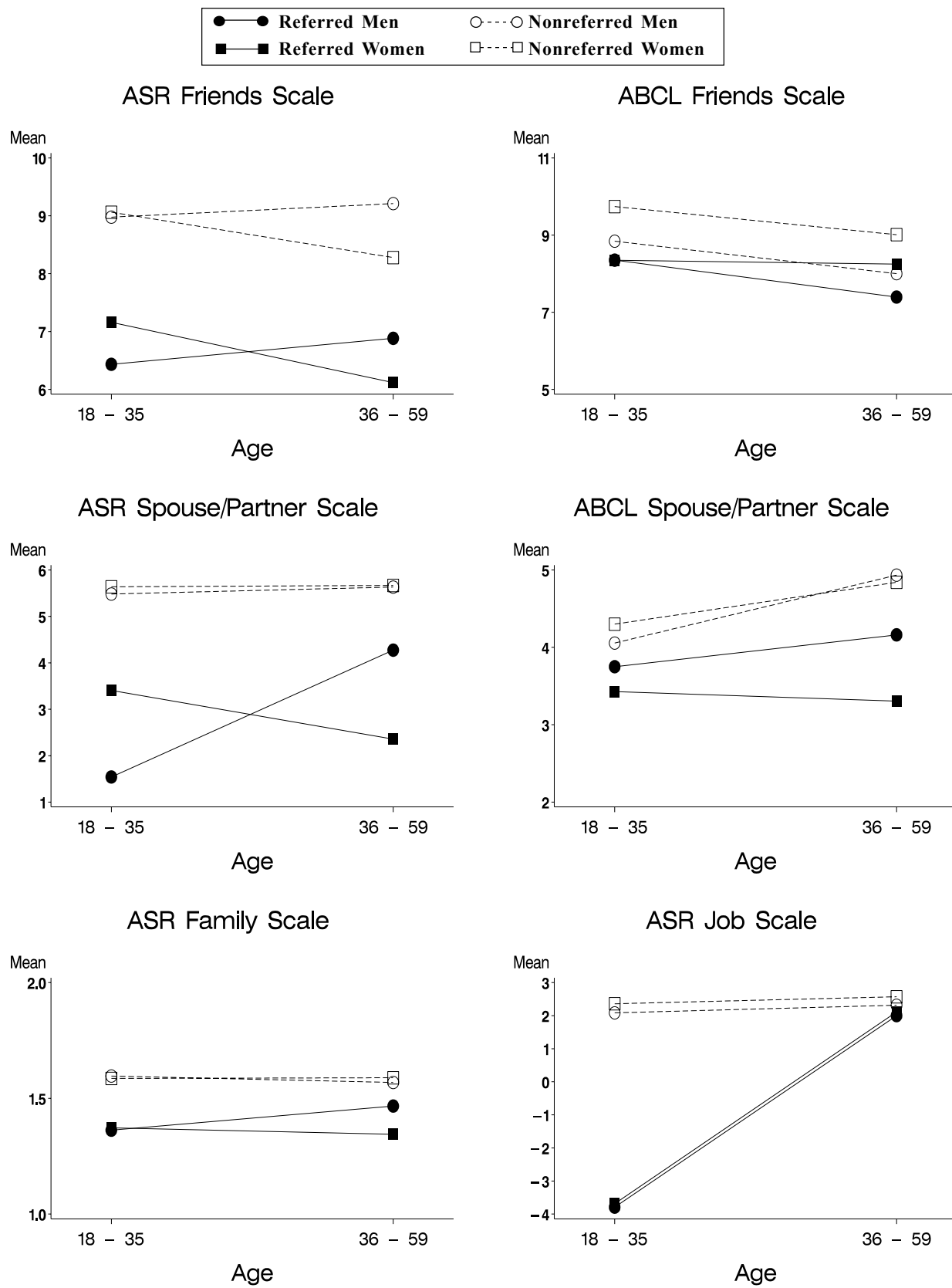
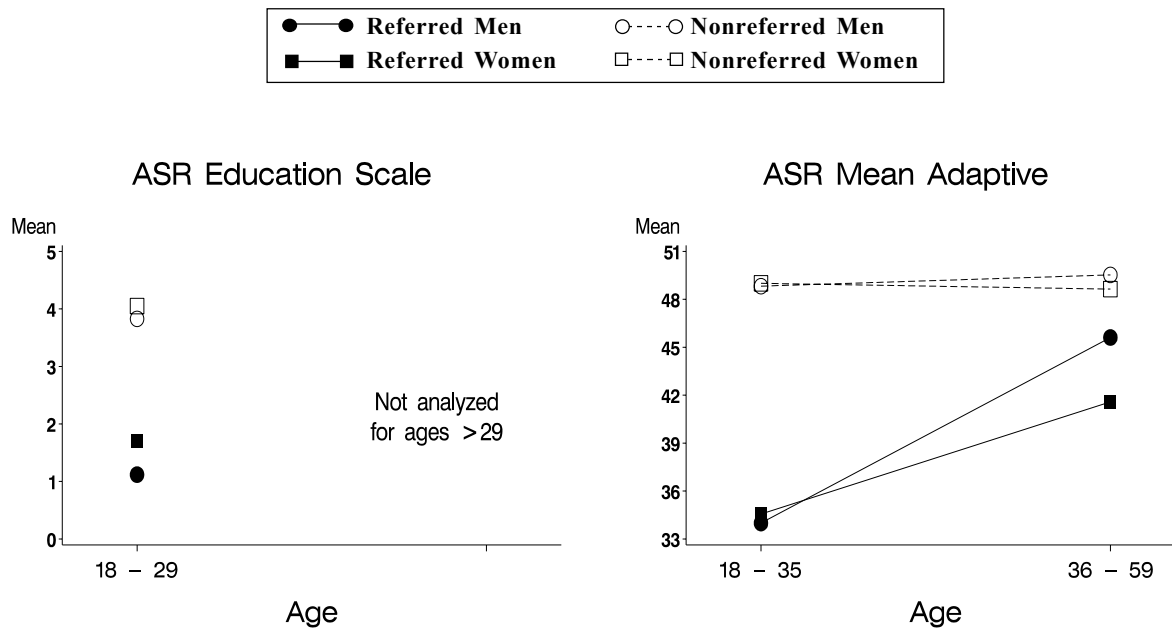**Figure 10-1. Mean scores for adaptive functioning scales.**

Figure 10-1 (cont.). Mean scores for adaptive functioning scales.

Table 10-3
Percent of Variance Accounted for by Significant (*p* <.01) Effects of Referral Status,
Age, and Gender on Problem Scale Scores in Multiple Regressions

Scales	Ref Status[a]		Age[b]		Gender[c]	
	ASR	ABCL	ASR	ABCL	ASR	ABCL
	N = 1,692	894	1,692	894	1,692	894
Substance Use						
Tobacco	3	6	1[O]	—	1[Md]	—
Alcohol	1[d]	2	—	—	1[M]	2[M]
Drugs	7	5	1[O]	—	1[M]	—
Mean	6	6	1[O]	—	2[M]	1[M]
Critical Items	20	8	1[Y]	1[Yd]	—	—
Empirically Based						
Anxious/Depressed	17	9	2[Y]	—	2[F]	1[F]
Withdrawn	8	7	—	—	—	—
Somatic Complaints	6	2	—	—	2[F]	1[Fd]
Thought Problems	3	1[d]	1[Y]	—	—	—
Attention Problems	15	8	2[Y]	3[Y]	1[F]	—
Aggressive Behavior	8	6	1[Y]	2[Y]	—	—
Rule-Breaking Behavior	15	11	1[Yd]	4[Y]	2[M]	1[M]
Intrusive	2	—	4[Y]	3[Y]	1[M]	2[M]
Internalizing	15	9	1[Yd]	—	1[F]	—
Externalizing	13	8	2[Y]	4[Y]	1[M]	—
Total Problems	13	8	2[Y]	2[Y]	—	—
DSM-Oriented						
Depressive Problems	22	9	1[Y]	—	2[F]	—
Anxiety Problems	2[d]	2	—	—	2[F]	—
Somatic Problems	4	1[d]	—	—	2[F]	—
Avoidant Personality Problems	11	7	1[Y]	—	<1[Fd]	—
AD/H Problems	10	6	1[Y]	3[Y]	—	2[M]
Inattention	10	7	1[Y]	2[Yd]	1[F]	—
Hyperactivity-Impulsivity	6	3	1[Y]	4[Y]	—	2[M]
Antisocial Personality Problems	15	9	2[Y]	4[Y]	1[M]	1[Md]

Note. Analyses were multiple linear regressions of raw scale scores on referral status, age, gender, and nonLatino white vs. other ethnicity. See text regarding other effects.

[a]All scale scores were significantly (*p* <.01) lower for nonreferred than referred adults except ABCL Intrusive.

[b]O = older adults scored higher; Y = younger adults scored higher.

[c]F = females scored higher; M = males scored higher.

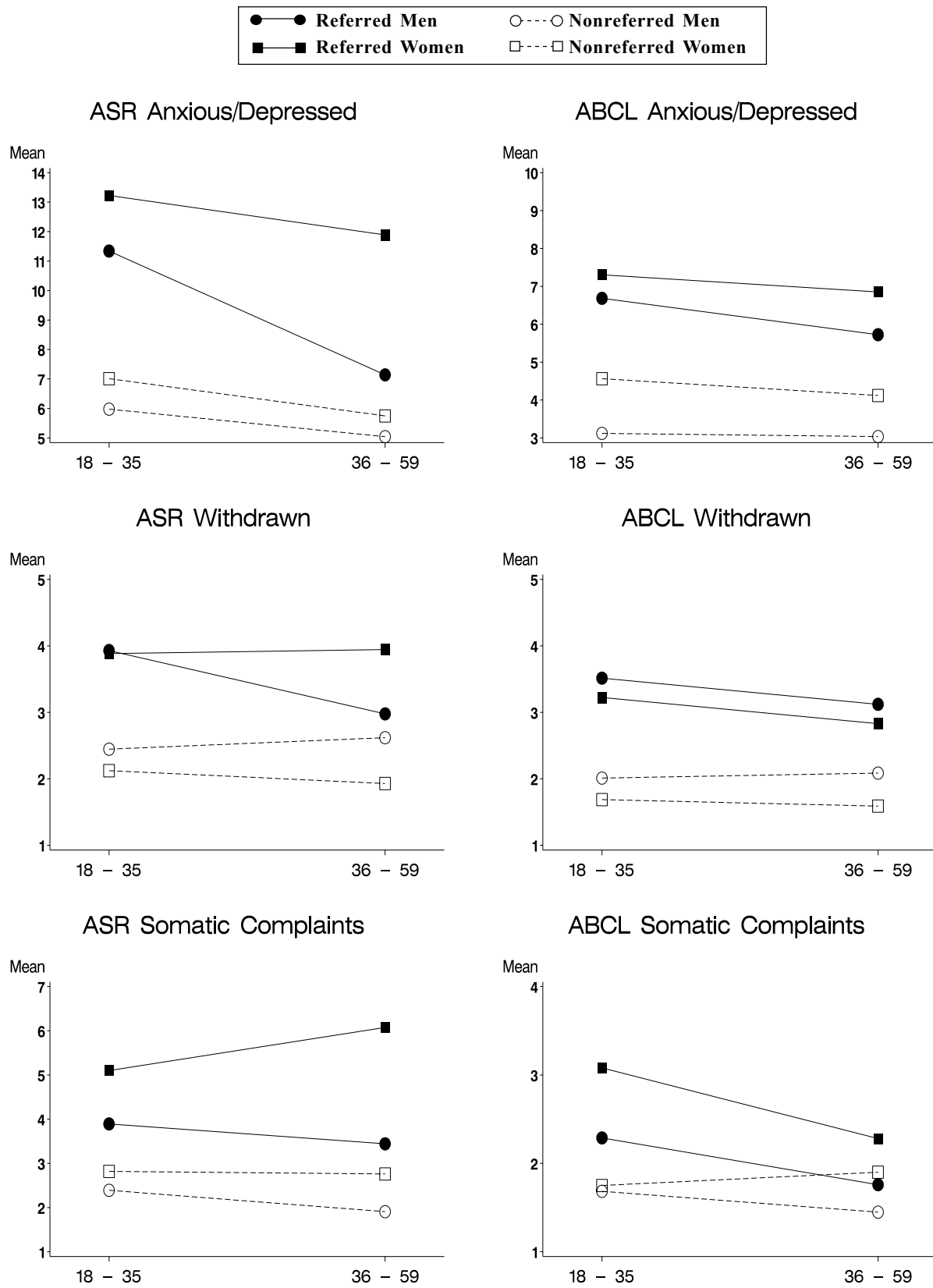[d]Not significant when corrected for number of analyses (Sakoda et al., 1954).

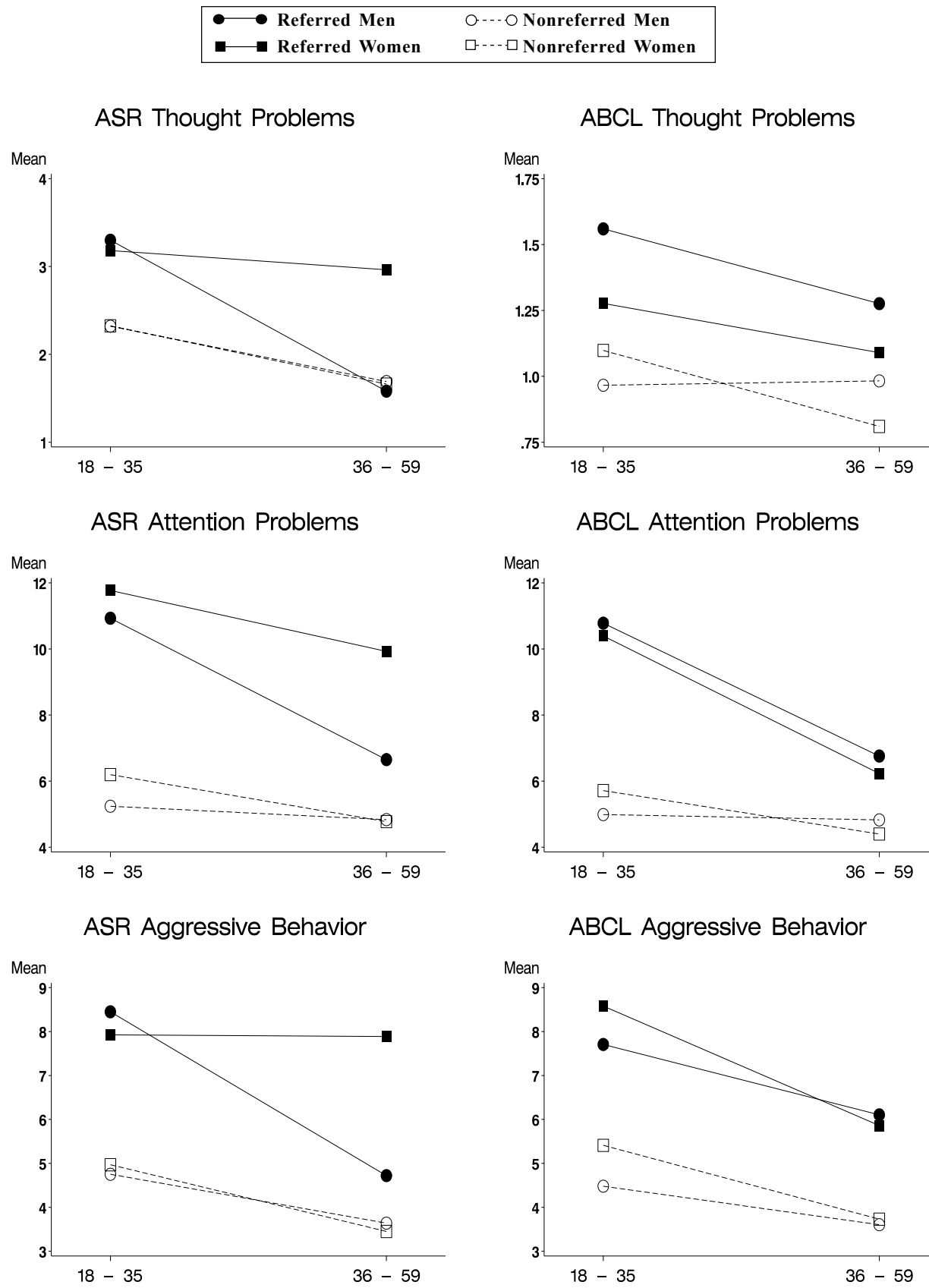**Figure 10-2. Mean scores for problem scales.**
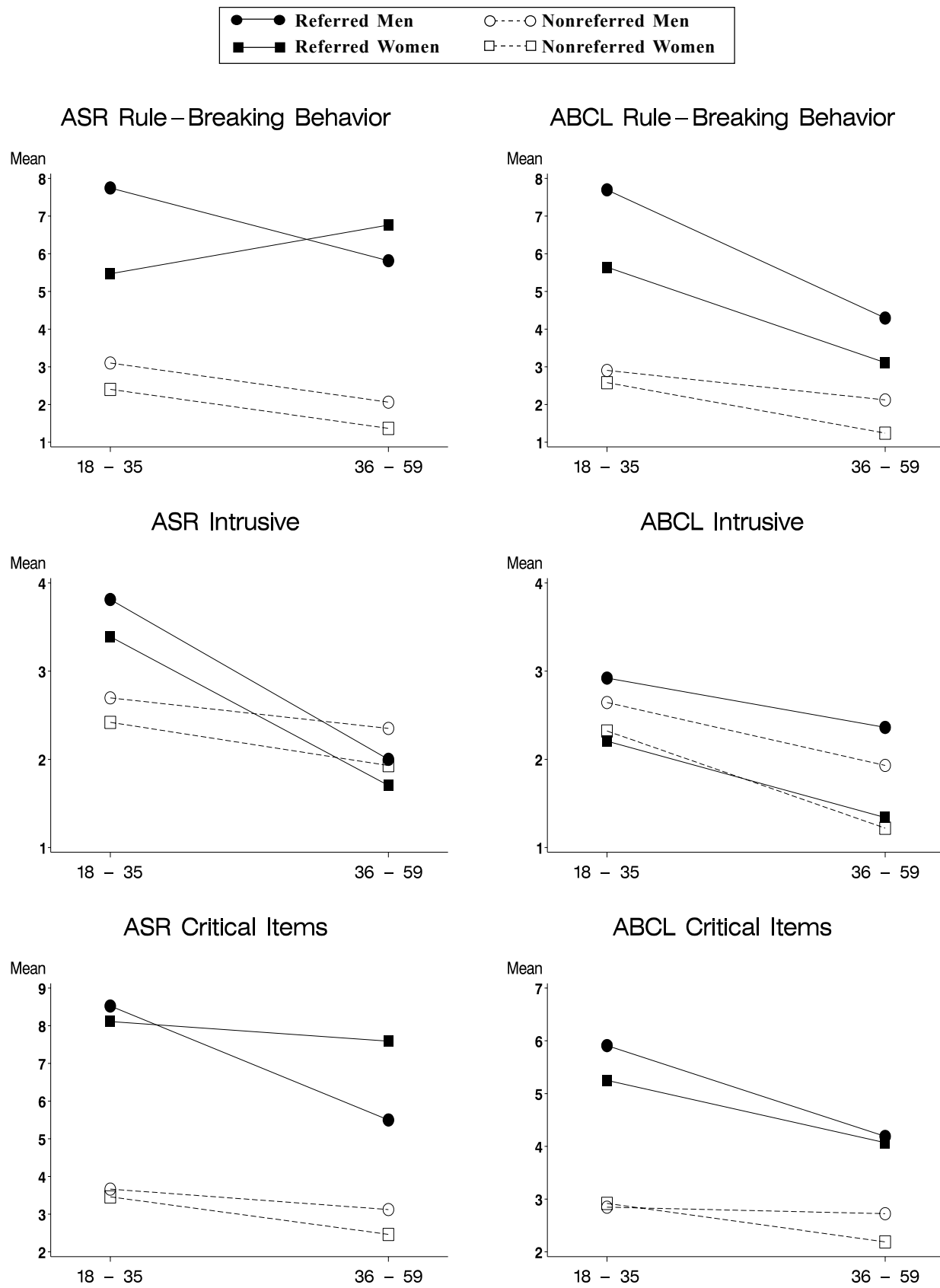
Figure 10-2 (cont.). Mean scores for problem scales.

Figure 10-2 (cont.). Mean scores for problem scales.
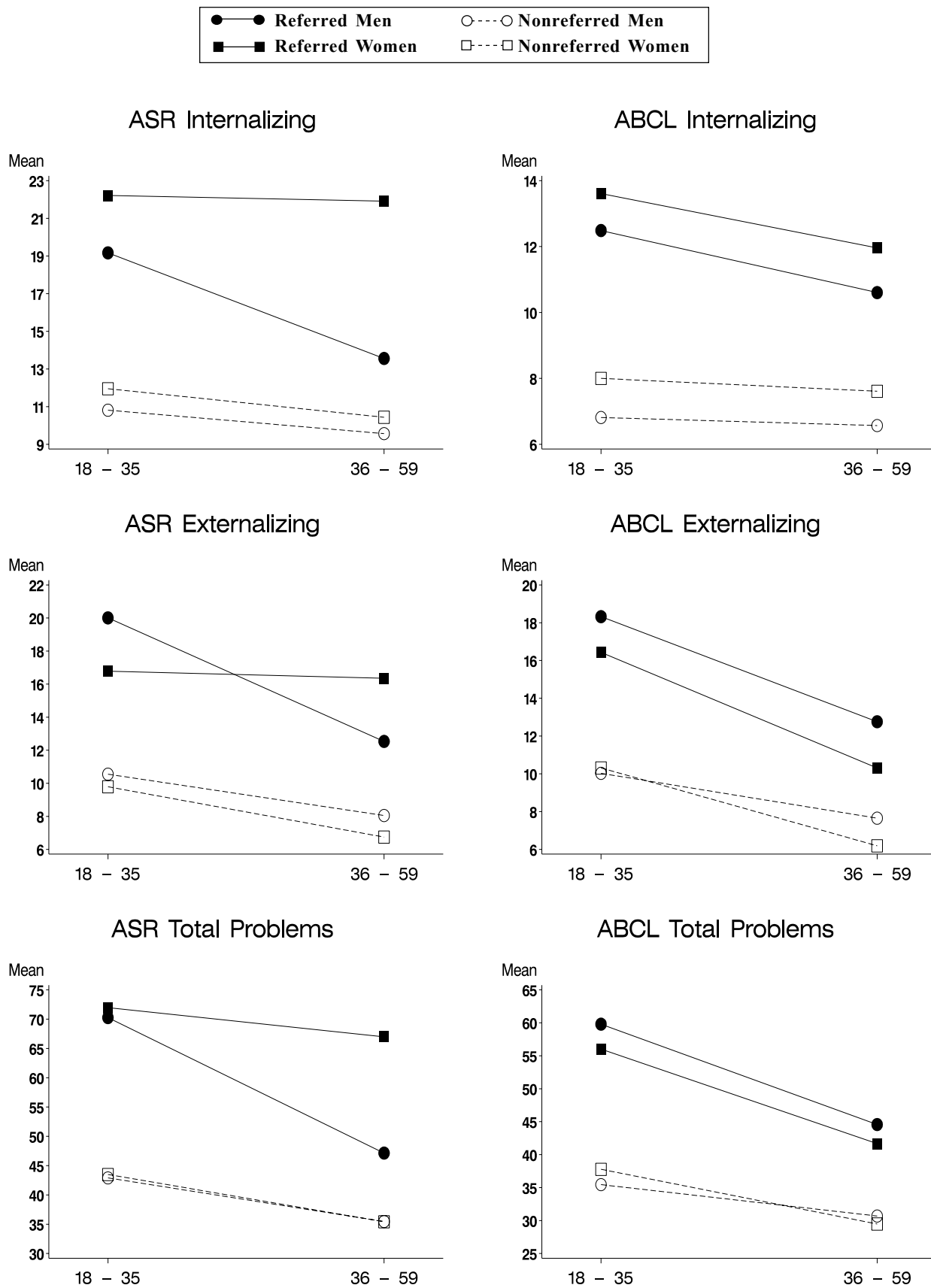
Figure 10-2 (cont.). Mean scores for problem scales.
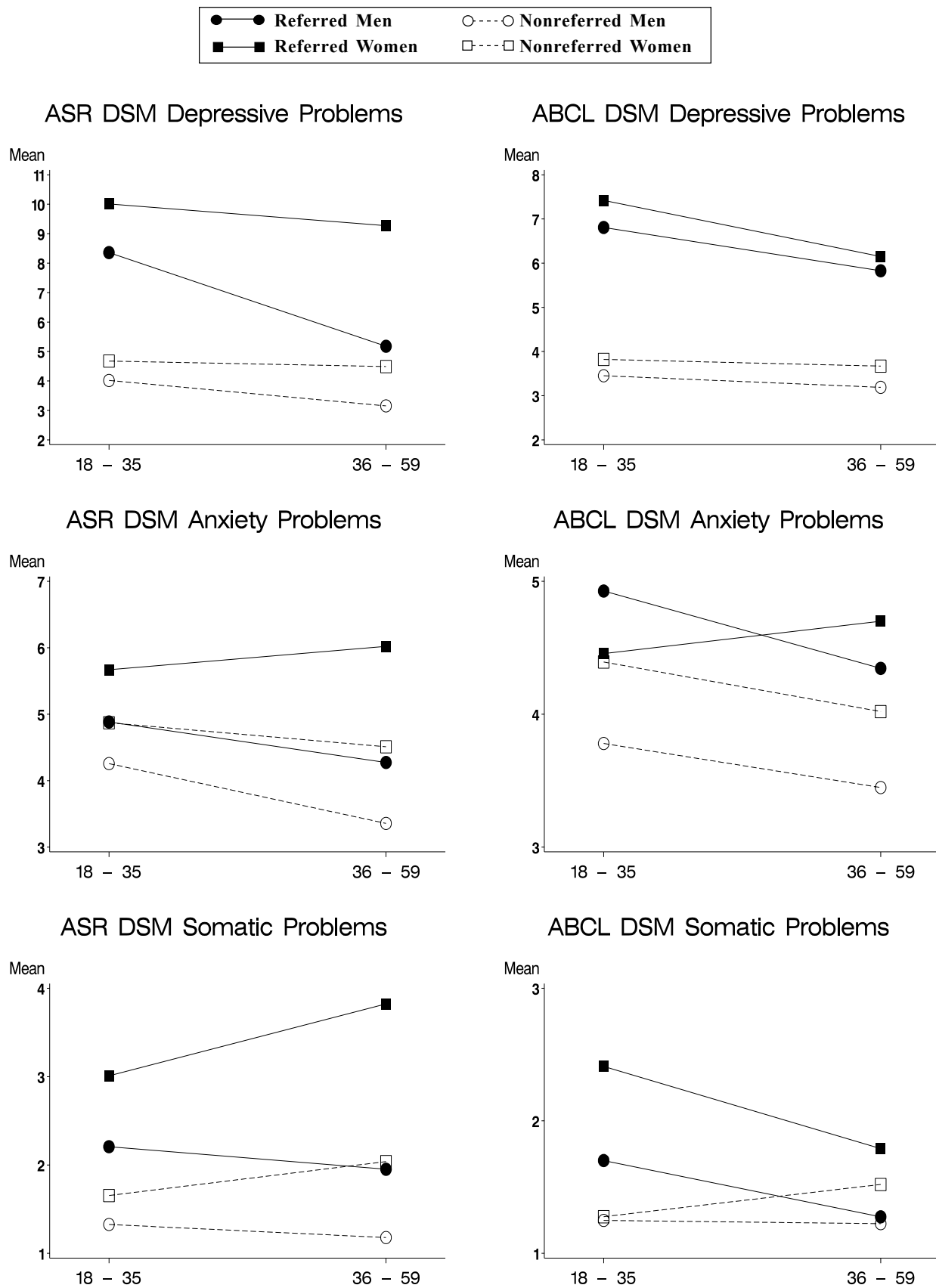
Figure 10-2 (cont.). Mean scores for problem scales.
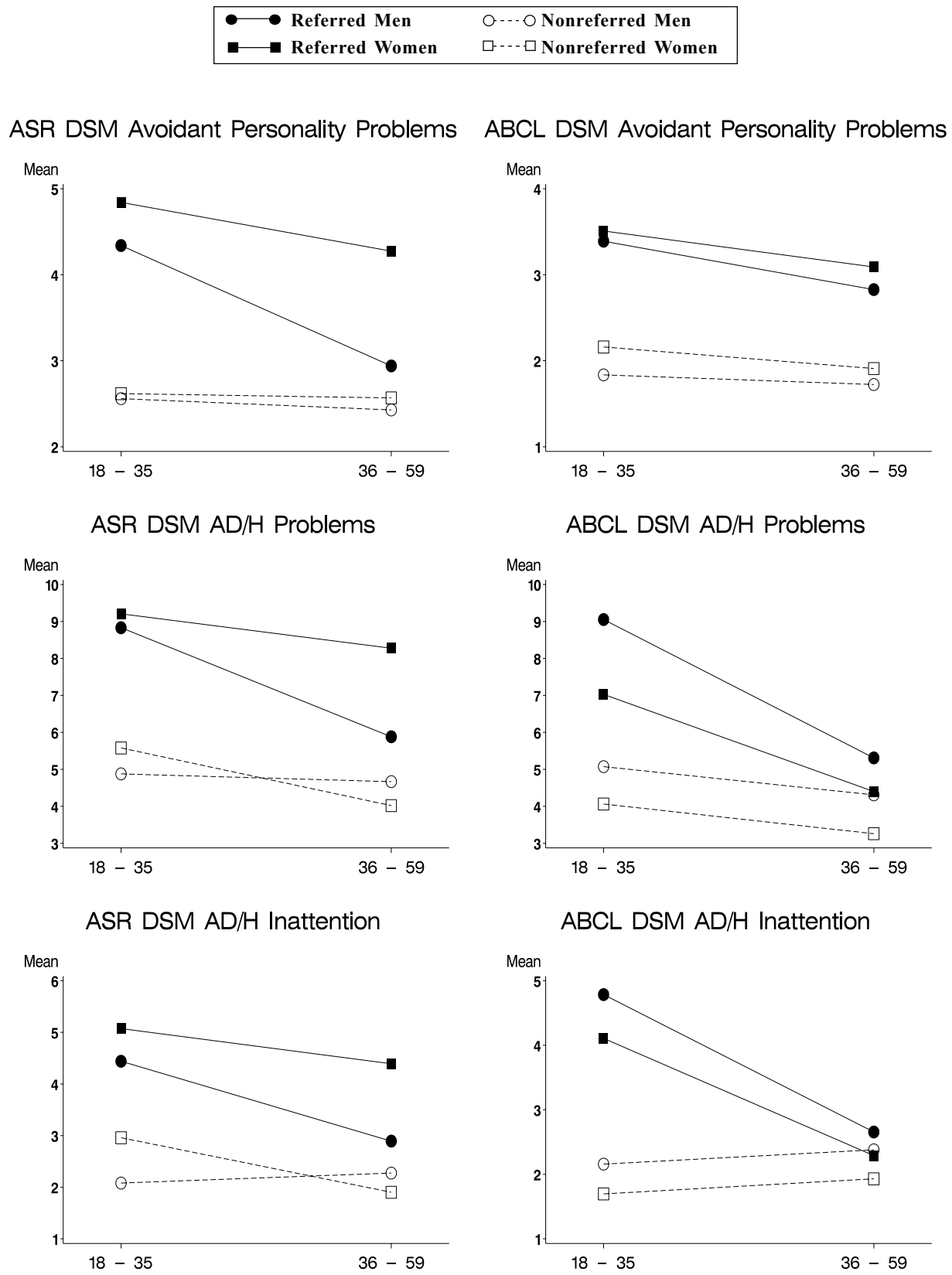
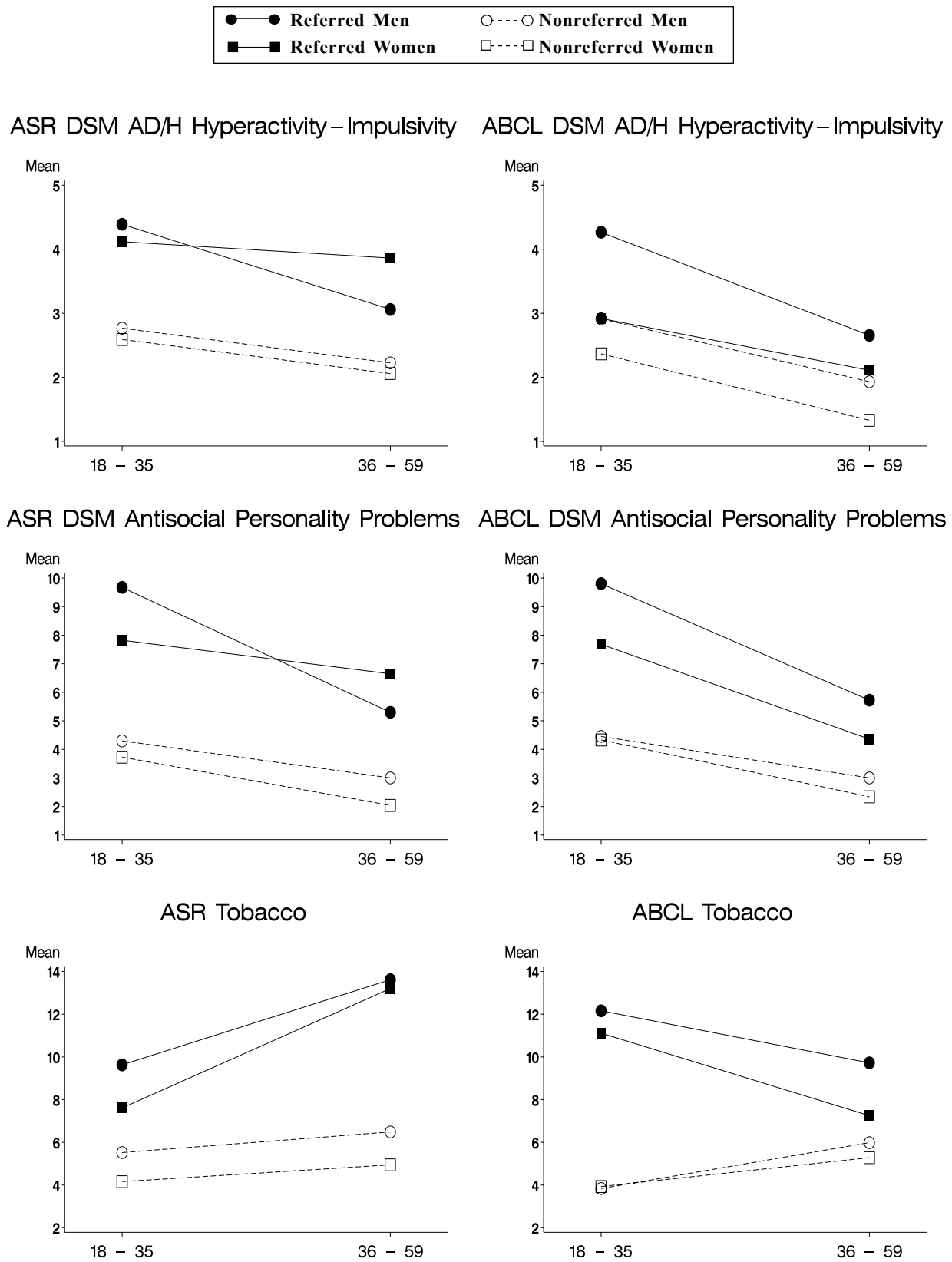**Figure 10-2 (cont.). Mean scores for problem scales.**

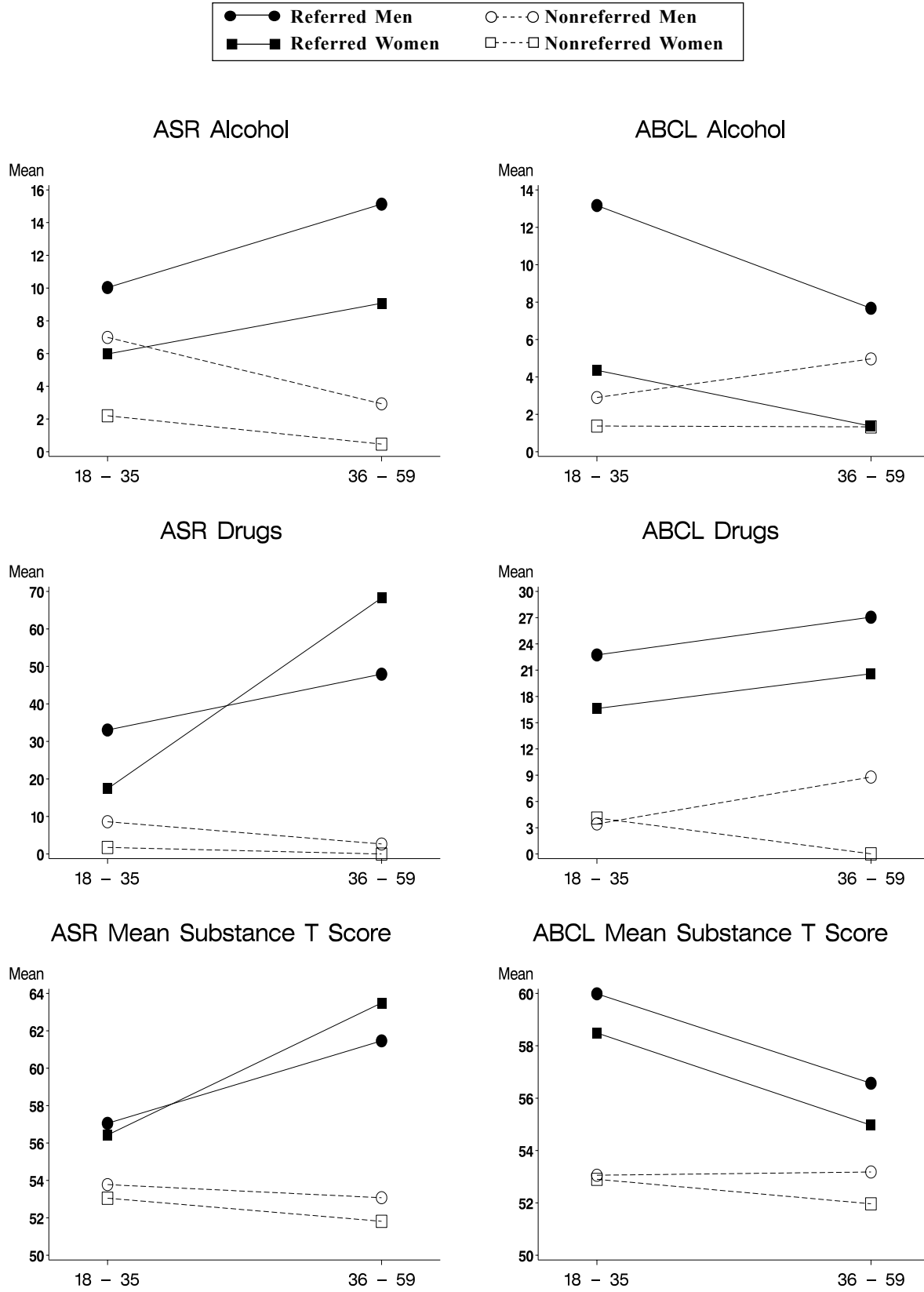**Figure 10-2 (cont.). Mean scores for problem scales.**

Figure 10-2 (cont.). Mean scores for problem scales.

egory, we have identified a borderline clinical range for each scale. The inclusion of a borderline category improves the basis for decisions about needs for help.

As an example, a scale score in the borderline range tells us that enough problems have been reported to be of concern but not so many that a person clearly needs professional help. If a person obtains one or more scale scores in the borderline range but none in the clinical range, we should consider options such as the following: *(a)* Obtain ratings from more informants to determine whether they view the person as being in the normal, borderline, or clinical range; *(b)* have the initial informants rate the person again after 2 to 3 months to see whether the borderline scores move into the normal or clinical range; *(c)* use additional assessment procedures to evaluate the kinds of problems comprising the scales that reached the borderline range. In other words, borderline scores can help users make more differentiated decisions than if all scores must be categorized as normal vs. clinical.

Continuous, quantitative scale scores afford greater statistical power than categorization of scores into a couple of levels, such as normal vs. borderline and clinical. Nevertheless, users may wish to distinguish dichotomously between normal and deviant scale scores. In the following sections, we report findings that indicate the degree to which dichotomous classifications of ASEBA scale scores according to the normal range vs. combined borderline and clinical ranges distinguish between demographically similar nonreferred vs. referred people. Because the borderline range encompasses scores that are high enough to be of concern, we have included it with the clinical range for our dichotomous comparisons of deviant scores with scores that are in the normal range.

Odds Ratios (ORs)

One approach to analyzing associations between two dichotomous ways of classifying people is by computing the *relative risk odds ratio* (OR; Fleiss, 1981), which is used in epidemiological research. The OR indicates the odds that people who have a particular risk factor also have a particular condition (usually a disorder), relative to the odds that people who lack the risk factor have the condition. The comparison

between outcome rates for those who do vs. do not have the risk factor is expressed as the ratio of the odds of having the outcome if the risk factor is present, to the odds of having the outcome if the risk factor is absent. For example, a study of relations between smoking (the risk factor) and lung cancer (the outcome) may yield a relative risk OR of 6. This means that people who smoke have 6 times greater odds of developing lung cancer than people who do not smoke.

We applied OR analyses to the relations between ASEBA scale scores and referral status as follows: For each ASEBA scale, we first classified people from our referred and nonreferred samples according to whether their scores were in the normal range or were deviant (i.e., were in the borderline or clinical range). Deviant scores were thus equivalent to a "risk factor" in epidemiological research, whereas referral vs. nonreferral was the outcome. We then computed the odds that people whose scores were deviant on a particular scale were from the referred sample, relative to the odds for people whose scores were not deviant on that scale.

The OR is a nonparametric statistic computed from a 2 x 2 table. For the analysis of each scale scored from each form, we therefore included both genders and all ages to provide a summary OR across all groups for whom the form was scored. The statistical significance of the OR is evaluated by computing confidence intervals.

Adaptive Functioning Scales. Table 10-4 displays the ORs for relations between deviant scores and referral status for the ASR and ABCL adaptive functioning scales. Table 10-4 also shows the percent of referred people whose scores were deviant according to the cutpoints for the normal vs. combined borderline and clinical ranges on each scale. For all ASR adaptive functioning scales, confidence intervals showed that the ORs were significantly ($p<.01$) greater than 1.0, while chi squares for all the scales showed that significantly more referred than nonreferred people obtained deviant scores ($p<.01$). The largest OR was 83 for the Job scale. For the two ABCL adaptive functioning scales, the confidence intervals and chi squares were not significant.

Table 10-4
Significant (*p*<.01) Odds Ratios and Percent of Adults Who Obtained
Deviant Scores on Adaptive Functioning Scales

| | Odds Ratios | | Percent Deviant | | | |
| | | | Referred | | Nonreferred | |
Scales	*ASR*	*ABCL*	*ASR*	*ABCL*	*ASR*	*ABCL*
	N = 1,692	296	846	148	846	148
Friends	3	—	19	12	7	6
Spouse/Partner	6	—	31	18	7	12
Family	3[b]	NA	12[b]	NA	5	NA
Job	83	NA	75	NA	3	NA
Education	8	NA	43	NA	8	NA
Mean Adaptive	34	NA	78	NA	10	NA

*Note. N*s vary becaue some scales were not relevant to some participants. Odds ratios indicate the odds that adults who obtained scores in the combined borderline and clinical ranges were referred for services. On all scales, the proportion of referred adults scoring in the clinical range significantly exceeded the proportion of nonreferred adults at *p*<.01 according to confidence intervals for odds ratios and chi squares for 2 x 2 tables, except ABCL Friends and Spouse/Partner scales.

[a]Deviant = combined borderline and clinical range.

[b]Not significant when corrected for number of analyses (Sakoda et al., 1954).

Problem Scales. Table 10-5 displays the ORs for relations between deviant scores and referral status for the problem scales and for combinations of cutpoints on the problem scales and adaptive scales. All ORs were significant at *p*<.01, except the ABCL DSM-oriented Anxiety Problems scale, which was *p*<.05, and the ABCL empirically based Thought Problems and Intrusive syndromes, which were both *p*>.05. The largest ORs were 9 for the ASR Critical Items scale and 8 for the ASR Attention Problems syndrome.

In addition to the ORs for individual problem scales, Table 10-5 also displays ORs for scores in the deviant range (combined borderline and clinical ranges) on the following combinations of scales: *(a)* ≥1 syndrome scale; *(b)* Internalizing and/or Externalizing; *(c)* Mean Adaptive and/or Total Problems; and *(d)* ≥1 DSM-oriented scale.

DISCRIMINANT ANALYSES

The foregoing sections dealt with the use of unweighted scale scores to discriminate between referred vs. nonreferred samples of people. Another approach is to use weighted combinations of scores to discriminate between groups. To test this approach, we performed stepwise discriminant analyses in which the criterion groups were the demographically similar samples of referred vs. nonreferred people.

We tested three sets of candidate predictors in the ASR and ABCL samples. The three sets of candidate predictors were: *(a)* the 8 syndrome scales; *(b)* the 6 DSM-oriented scales; and *(c)* all problem items on a form. We did not use the adaptive functioning items because they had not been assessed for all the participants. An alpha criterion of *p*<.05 was used for entry of candidate predictors into the discriminant functions.

Discriminant analyses selectively weight candidate predictors to maximize their collective associations with the particular criterion groups being analyzed, using characteristics of the sample that may differ from other samples. To avoid overestimating the accuracy of the classification obtained by discriminant analyses, it is necessary to correct for the "shrinkage" in associations that may occur when discriminant weights derived in one sample are applied to a new sample.

Table 10-5

**Significant (*p*<.01) Odds Ratios and Percent of Referred and Nonreferred Adults
Who Obtained Deviant Scores on Problem Scales**

| | Odds Ratios | | Percent Deviant[a] | | | |
| | | | Referred | | Nonreferred | |
Scales	ASR	ABCL	ASR	ABCL	ASR	ABCL
Empirically Based	*N* = 1,692	894	846	447	846	447
Anxious/Depressed	6	4	29	23	7	7
Withdrawn	4	3	22	16	7	6
Somatic Complaints	3	2[b]	23[b]	13[b]	8	6
Thought Problems	2[b]	—	18	11	8	8
Attention Problems	8	3	35	20	7	8
Aggressive Behavior	3	2	22	19	8	9
Rule-Breaking Behavior	6	5	33	25	7	7
Intrusive	2[b]	—	14[b]	11	8	9
Internalizing	5	3	49	35	16	15
Externalizing	4	3	43	40	17	19
Total Problems	4	3	47	37	17	17
≥ 1 syndrome in deviant range	3	3	59	53	32	28
Int and/or Ext in deviant range	4	3	57	52	25	26
Mean Adaptive and/or Total						
Problems in deviant range	4	NA	57	NA	25	NA
Critical Items	9	4	41	23	7	8
DSM-Oriented						
Depressive Problems	7	4	38	24	8	8
Anxiety Problems	3	2[b]	14	13[b]	6	7
Somatic Problems	3	2[b]	20	15	8	9
Avoidant Personality Problems	4	3	22	20	6	8
AD/H Problems	6	3	29	18	6	8
Inattention	6	3	28	19	6	8
Hyperactivity-Impulsivity	4	2[b]	22	14[b]	7	8
Antisocial Personality Problems	6	5	32	27	7	7
≥ 1 DSM scale in deviant range	4	3	59	52	27	26
Substance Use						
Tobacco	3	4	20	32	8	11
Alcohol	2[b]	3	14[b]	17	8	7
Drug	6	7	28	24	6	7
Mean Substance	5	5	29	27	7	7

Note. Odds ratios indicate the odds that referred adults obtained scores in the combined borderline and clinical ranges. On all scales, the proportion of referred adults scoring in the clinical range significantly exceeded the proportion of nonreferred adults at *p* <.01 according to confidence intervals for odds ratios and chi squares for 2 x 2 tables except ABCL DSM-oriented Anxiety Problems (*p*<.05), and empirically based Thought Problems and Intrusive (both *p*>.05).

[a]Deviant = combined borderline and clinical range.

[b]Not significant when corrected for number of analyses (Sakoda et al., 1954).

Cross-Validated Correction for Shrinkage

To correct for shrinkage, we used a "jackknife" procedure whereby the discriminant function for each sample was computed multiple times with a different person held out of the sample each time (SAS Institute, 1999). Each discriminant function was then cross-validated by testing the accuracy of its prediction for each of the "hold-out" people. Finally, the percentage of correct predictions was computed across all the hold-out people. It is these cross-validated predictions that we will present.

Cross-Validated Percent of People Correctly Classified

Table 10-6 displays the cross-validated percent of people who were correctly classified by the discriminant analyses using the three different sets of candidate predictors for each instrument. As you can see in Table 10-6, the three sets of predictors achieved total accuracies ranging from 71% (syndromes) to 87% (problem items) for the ASR. For the ABCL, the total accuracies were quite similar for the three sets of predictors, ranging from 65% for the problem items to 68% for the syndrome scales and for the DSM-oriented scales. By looking again at Table 10-6, you can see that the ASR problem items produced the best results for each of the four classification parameters, as follows: True positives (sensitivity) = 80%; true negatives (specificity) = 95%; false positives = 20%; false negatives = 5%.

Results for Specific Scales. In the discriminant analyses that used the eight syndrome scales as candidate predictors, the following syndromes survived as significant predictors for both the ASR and ABCL: Anxious/Depressed; Thought Problems; and Rule-Breaking Behavior. For the ASR, the Attention Problems syndrome and, for the ABCL, the Intrusive syndrome also survived. Among the syndromes that survived as significant predictors for both the ASR and ABCL, one was from the Internalizing grouping, one was from the Externalizing grouping, and one was a syndrome that was not strongly associated with either the Internalizing or Externalizing grouping. These findings indicate that no one pattern of problems outweighs all others as a significant predictor of need for services.

From the six DSM-oriented scales, the following four survived as significant predictors for both the ASR and ABCL: Depressive Problems; Anxiety Problems; Avoidant Personality Problems; and Antisocial Personality Problems. Based on both the syndromes and the DSM-oriented scales, it is thus evident that diverse sets of problems contribute significantly to discriminating between people who are deemed to need professional help and those who are not.

Results for Specific Problem Items. For the ASR, 33 items survived as significant predictors, while for the ABCL, 14 survived. In both the ASR and ABCL analyses, the first item to be entered was *103. Unhappy, sad, or depressed.* Item 103 has been repeatedly found to be an especially powerful discriminator between referred vs. nonreferred children and adolescents according to different informants in multiple samples (Achenbach, 1991b,d; Achenbach & Edelbrock, 1983, 1986, 1987; Achenbach & Rescorla, 2001; Verhulst, Akkerhuis, & Althaus, 1985). Even when entered with a very large number of other predictors, this item often obtains considerably larger discriminant function coefficients than the other significant predictors. As shown in Chapter 11, ANCOVAs of all problem items revealed that item *103* was more strongly associated with referral status than any other problem item on the ASR (ES = 27%) and ABCL (ES = 11%).

For both the ASR and ABCL, the second item to be entered was *6. Uses drugs (other than alcohol or nicotine) for nonmedical purposes.* The survival of this item as a strong predictor on both the ASR and ABCL indicates that drug use is likely to be associated with a substantial proportion of referrals for services.

The only other item to survive as a significant predictor in the discriminant functions for both the ASR and ABCL was *47. Lacks self-confidence.*

In summary, the discriminant analyses achieved the best cross-validated accuracy of 87% of participants

Table 10-6
Cross-Validated Percent of Adults Correctly Classified as Referred
vs. Nonreferred by Discriminant Analyses

| | *ASR* | | | | | *ABCL* | | | | |
| | *% Correctly* | | | | | *% Correctly* | | | | |
Candidate Predictors	*Classified*	*TP*	*TN*	*FP*	*FN*	*Classified*	*TP*	*TN*	*FP*	*FN*
8 Syndromes	71	65	78	22	35	68	56	80	20	44
6 DSM- oriented scales	77	72	82	18	28	68	57	79	21	43
All problem items	87	80	95	5	20	65	55	76	24	45

Note. TP = true positives (sensitivity); TN = true negatives (specificity); FP = false positives; FN = false negatives.

correctly classified when selecting from among all the problem items on the ASR. The priority of item *103. Unhappy, sad, or depressed* as a predictor on both the ASR and the ABCL and its strong discriminative power in analyses of numerous samples of children and adolescents attest to its association with diverse conditions that warrant professional help across a wide age span.

PROBABILITY OF PARTICULAR TOTAL PROBLEMS SCORES BEING FROM THE REFERRED VS. NONREFERRED SAMPLES

To provide further perspectives on relations between ASEBA scores and referral status, Table 10-7 displays the probabilities that particular *T* scores were from referred samples rather than from demographically similar nonreferred samples. The probabilities were determined by tabulating the proportion of people from our matched referred and nonreferred samples within each of the *T* score intervals shown in Table 10-7. We used *T* scores in order to provide a uniform metric across all gender/age groups on the ASR and ABCL.

The probability that a score was from the referred sample increased fairly steadily as the Total Problems scores increased. Users can consult Table 10-7 to estimate the probability that particular Total Problems scores represent deviance severe enough to warrant concern.

CONSTRUCT VALIDITY OF ASEBA SCALES

According to a dictionary definition, a *construct* is "an object of thought constituted by the ordering or systematic uniting of experiential elements" (Gove, 1971, p. 489). ASEBA scales can be viewed as representing constructs that have been derived by systematically ordering scores on the items of the ASEBA forms, which tap people's experience pertaining to the individual being assessed.

Construct validity concerns evidence that supports hypothesized variables (*hypothetical constructs*) for which there are no definitive criterion measures. A primary reason for developing ASEBA instruments was to derive syndromal constructs that embody patterns of problems that occur together. Studies of ASEBA child and adolescent syndromes have revealed numerous correlates and considerable developmental stability for the syndromes (evidence has been reviewed by Achenbach & Rescorla, 2000, 2001).

The correlates and developmental courses of the different syndromes indicate that they reflect important differences in patterns of child and adolescent functioning. The validity of constructs such as the ASEBA syndromes is supported by the accumulation of evidence for systematic relations between measures of the constructs and other variables. These systematic relations are called *nomological* (i.e., lawful) *networks* (Cronbach & Meehl, 1955).

Table 10-7
Probability of Total Problems *T* Scores Being from Referred Samples

Total Problems T Scores	ASR	ABCL
	N = 1,692	*N* = 894
0-35	.27	.19
36-39	.20	.33
40-43	.30	.15
44-47	.37	.35
48-51	.40	.53
52-55	.46	.45
56-59	.50	.56
60[a]-63	.60	.67
64-67	.69	.68
68-100	.87	.72

Note. Samples were demographically similar referred and nonreferred adults.
[a]*T* scores ≥ 60 are in the combined borderline and clinical range.

Each ASEBA syndrome can be viewed in statistical terms as representing a *latent variable* derived by factor analyzing ASEBA items. The versions of a syndrome derived from separate factor analyses of the ASR and ABCL provide different ways of operationalizing the construct represented by the syndrome. Furthermore, the versions of a syndrome scored from ASR and ABCL ratings provide different quantitative measures of the latent variables represented by the syndromes.

People differ in their knowledge of an individual's functioning, in their roles in relation to the individual being assessed, in what they remember, and in personal characteristics that can affect their ratings. Consequently, the correlations among ratings by different respondents, especially those playing different roles with respect to the individual they rate, may be modest, as shown in Chapter 9. Nevertheless, the test-retest reliability of ASR and ABCL ratings is good, as documented in Chapter 9. Furthermore, the content and criterion-related validity of ratings by different informants has been documented in the preceding sections of this chapter. The findings thus indicate that different informants can contribute to the assessment process.

Assessment of the syndromal constructs via data from multiple sources is consistent with the way in which psychological constructs are conceptualized and evaluated. Because psychological constructs involve inferences about latent variables that are not directly observable, their validity must be evaluated in terms of various kinds of indirect evidence relevant to their validity. The *Bibliography of Published Studies Using ASEBA Instruments* (Bérubé & Achenbach, 2003) lists some 4,500 published studies of ASEBA instruments. Many of the studies provide evidence for the construct validity of ASEBA scales in terms of significant associations with other variables, prediction and evaluation of outcomes, and consistency with theoretical formulations. In the following sections, we summarize several kinds of support for the construct validity of ASEBA adult scales. Although some of the findings were obtained with earlier versions of the scales, the earlier versions correlate highly with the 2003 versions, as documented in Chapter 12.

Correlations of ASR Scales with SCL-90-R Scales

The Symptom Checklist-90-Revised (SCL-90-R; Derogatis, 1994) includes 90 self-report problem items

that are rated on 5-point scales ranging from *0 = Not at all* to *4 = Extremely.* Respondents rate each item according to "how much the problem has distressed or bothered you during the past 7 days including today." Items were chosen to assess the following nine symptom constructs: *Somatization, Obsessive-Compulsive, Interpersonal Sensitivity, Depression, Anxiety, Hostility, Phobic Anxiety, Paranoid Ideation,* and *Psychoticism.* Each construct is scored by summing the 0 to 4 ratings on the items measuring the construct. The items of the SCL-90-R are also scored in terms of the following three global indices: The Global Severity Index (GSI), which is computed by summing the ratings on all items and dividing by the number of items that were rated; the Positive Symptom Total (PST), which is the number of items that were endorsed; and the Positive Symptom Distress Index, which is calculated by dividing the sum of all 1-4 ratings by the PST.

Like the ASR, the SCL-90-R can be self-administered under many conditions in about 15 to 20 minutes. Unlike the ASR, the SCL-90-R does not have adaptive functioning items, open-ended items that invite respondents to provide details, substance use scales, DSM-oriented scales, or the kinds of items that comprise the ASR's Rule-Breaking Behavior and Intrusive syndromes. However, because the SCL-90-R and ASR are apt to be candidates for use under similar clinical and research conditions, it is helpful to know the degree to which their scales tap similar constructs. Table 10-8 displays Pearson *r*s between ASR problem scales and SCL-90-R scales scored from forms completed by 66 clients of a research clinic for treatment of substance abuse in Vermont. (The SCL-90-R PST global index is not shown, because its function is similar to that of the other SCL-90-R global indices, both of which correlated higher than the PST with the ASR scales.)

There were significant correlations between every ASR problem scale and every SCL-90-R scale shown in Table 10-8. Most of the correlations met Cohen's (1988) criteria for large ES ($r \geq .50$). Twenty of the *r*s were $\geq .70$. There were thus strong and pervasive associations between scores on the two instruments. Among the ASR syndromes, the Anxious/Depressed syndrome had its highest *r*s (.69 to .78) with the SCL-

90-R Obsessive-Compulsive, Interpersonal Sensitivity, Depression, and Anxiety dimensions. The ASR Somatic Complaints syndrome had its highest *r* (.66) with the Somatization dimension. The ASR Thought Problems syndrome had its highest *r*s (both .53) with the Obsessive-Compulsive and Psychoticism dimensions. The ASR Attention Problems syndrome had its highest *r* (.70) with the Obsessive-Compulsive dimension. And the Aggressive Behavior syndrome had its highest *r* (.69) with the Hostility dimension. Reflecting the lack of corresponding SCL-90-R items, the highest *r*s of the ASR Rule-Breaking Behavior and Intrusive syndromes reached only .43 and .41, respectively, with an SCL-90-R dimension. However, the Rule-Breaking Behavior syndrome correlated .49 with the SCL-90-R Positive Symptom Total. Reflecting the predominance of internalizing symptoms on the SCL-90-R, the ASR Internalizing scale had *r*s in the .60s and .70s with 8 of the 9 SCL-90-R symptom dimensions, the Global Severity Index, and the Positive Symptom Total.

Among the ASR DSM-oriented scales, *r*s in the .70s were found for the Depressive Problems scale with the SCL-90-R Depression dimension, for the Avoidant Personality Problems scale with the SCL-90-R Interpersonal Sensitivity dimension, and for the Attention Deficit/Hyperactivity Problems scale with the SCL-90-R Obsessive-Compulsive scale. The fact that every SCL-90-R scale correlated $\geq .62$ with one or more ASR scales indicates that the types of problems assessed by the SCL-90-R are also assessed by the ASR.

Longitudinal Relations Between Pre-Adult Scales and Adult Scales

American National Sample: 3-Year Predictive Relations. Longitudinal studies have demonstrated significant predictive correlations from scores on the YSR and CBCL to scores on YASR and YABCL scales that correlate highly with the 2003 adult scales. In an American national sample, 3-year longitudinal *r*s from syndrome scores on YSRs completed by adolescents to syndrome scores on YASRs completed by the same participants 3 years later ranged from .34 for 367 females on the Somatic Complaints syndrome to .54 for 340 males on the Anxious/Depressed syndrome (Achenbach et al., 1995c).

Table 10-8

Significant ($p<.01$) Correlations of ASR Scales with SCL-90-R Scales

ASR Scales	Somati-zation	Obsess.-Compuls.	Interper. Sensitivity	Depres-sion	Anxiety	Hostility	Phobic Anx.	Paranoid Ideation	Psychot-icism	Global Severity	Symptom Total
				SCL-90-R Scales							
Empirically Based											
Anx./Dep.	.57	.71	.78	.76	.69	.47	.65	.62	.68	.76	.68
Withdrawn	.42	.60	.60	.46	.44	—	.46	.49	.57	.55	.57
Somatic	.66	.62	.57	.62	.58	—	.58	.48	.53	.66	.60
Thought	.49	.53	.49	.41	.49	.35[a]	.52	.40	.53	.53	.48
Attention	.46	.70	.60	.62	.58	.42	.61	.48	.60	.65	.64
Aggressive	.41	.58	.67	.63	.53	.69	.45	.50	.55	.63	.57
Rule-Break.	—	.43	—	.36[a]	—	.41	—	—	.39	.39	.49
Intrusive	—	.35[a]	—	—	—	.41	—	—	—	—	—
Internal.	.64	.75	.77	.75	.69	.46	.67	.62	.70	.78	.72
External.	.37[a]	.58	.57	.55	.49	.58	.41	.40	.51	.57	.57
Total	.56	.74	.72	.70	.65	.53	.62	.57	.67	.74	.70
DSM-Oriented											
Depressive	.54	.69	.69	.77	.65	.42	.61	.54	.61	.73	.62
Anxiety	.54	.62	.66	.64	.64	—	.60	.53	.59	.67	.62
Somatic	.63	.59	.53	.57	.53	.38[a]	.53	.45	.50	.62	.56
Avoidant	.51	.63	.70	.53	.52	.39[a]	.53	.56	.60	.63	.66
AD/H	.52	.71	.61	.58	.59	.45	.59	.43	.58	.65	.63
Antisocial	—	.55	.52	.49	.45	.52	.40	.41	.52	.53	.57

Note. Participants were 66 adults enrolled in a substance abuse treatment program. SCL-90-R Symptom Distress Index is not listed, because its significant correlations with ASR scales were all lower than the corresponding SCL-90-R Severity Index and Positive Symptom Total correlations with ASR scales.

[a]Not significant when corrected for the number of analyses (Sakoda et al., 1954).

Also in the American national sample, scores on CBCL syndromes yielded 3-year longitudinal *r*s with YABCL syndromes that ranged from .40 for the Somatic Complaints syndrome among 377 females to .67 for the Anxious/Depressed syndrome among 366 males. Including Internalizing, Externalizing, and Total Problems, the mean 3-year *r* from the CBCL scores to the YABCL scores was .59, which was very similar to the mean *r* of .60 between YABCL scales rated twice over 44 months by parents in the same national sample, as was shown in Table 9-2. These *r*s were somewhat higher than the YSR to YASR *r*s, but the pattern of lowest and highest *r*s was the same as in the self-ratings. The overall mean *r* of .58 between all the YASR scales rated twice over 39 months shown in Table 9-2 was also higher than the mean YSR to YASR *r* of .48. The younger ages of the participants when they completed the YSR may explain the somewhat lower *r*s than were found for the 39-month stability of the YASR. In any event, the similarity of the predictive *r*s between ratings of adolescents on the YSR and CBCL to young adult YASR/YABCL ratings at two points in adulthood indicate that the YASR/YABCL scales tapped the same constructs as the YSR and CBCL scales. An additional finding was that certain scores had strong associations with adult signs of disturbance, including police contacts, suicidal behavior, and drug use (Achenbach, Howell, McConaughy, & Stanger, 1998).

Dutch General Population Sample: 4-Year Predictive Relations. As part of a longitudinal study of randomly selected Dutch children, 364 adolescents completed YSRs at ages 15 to 18 (Ferdinand, Verhulst, & Wiznitzer, 1995). Two years later, they completed YASRs, and did so again 2 years after that. Over the 4-year interval from the YSR assessment to the second YASR assessment, the mean *r* between YSR and YASR Total Problems scores was .49, compared to a mean *r* of .55 found over a 3-year interval in an American national sample (Achenbach et al., 1995c). Allowing for the 1-year longer interval in the Dutch sample, the predictive relations from YSR to YASR Total Problems scores were thus very similar for American and Dutch young people.

Ferdinand et al. also classified participants according to whether they scored ≥90th percentile or ≤50th percentile on the YSR Total Problems, Internalizing, or Externalizing scales. The researchers then compared outcomes on the YASR 4 years later for participants who had initially been classified as high vs. low scorers on the YSR. Classification as high vs. low on the initial YSR significantly predicted classification as high vs. low on the YASR 4 years later for Total Problems, Internalizing, and Externalizing.

Dutch General Population Sample: 8-Year Predictive Relations. In the same Dutch general population sample, Ferdinand and Verhulst (1995) found that parents' CBCL ratings of 459 13- to 16-year-olds significantly predicted high problem scores on YASRs completed by the participants themselves 8 years later, when they were 21 to 24 years old. Among the eight cross-informant syndromes scored from the CBCL, the Anxious/Depressed and Somatic Complaints syndromes were especially good 8-year predictors of YASR Total Problems scores. The specificity of prediction from CBCL syndromes to YASR syndromes was greatest for the Withdrawn, Somatic Complaints, Anxious/Depressed, and Aggressive Behavior syndromes. The findings thus support the construct validity of these syndromes across informants over 8 years of transition from adolescence to adulthood, i.e., parents' CBCL ratings predicted YASR ratings by their adult offspring 8 years later.

Dutch General Population Sample: 10-Year Predictive Relations. In a 10-year reassessment of the same Dutch general population sample, Hofstra, van der Ende, and Verhulst (2001) tested the ability of YSRs administered at ages 11 to 19 years to predict YASR scores and DSM diagnoses obtained at ages 21 to 30 years. Linear regression analyses were used to test the prediction of scores on each YASR adult syndrome from scores on all YSR syndromes, plus age when the YSR was completed and SES, separately for each gender. In 14 of the 16 regression analyses (8 syndromes separately for each gender), the YASR syndromes were predicted most strongly by their counterpart YSR syndromes. (Because the YSR Aggressive Behavior syndrome has been found to split into the adult Aggressive Behavior and Intrusive syndromes,

the YSR Aggressive syndrome is viewed as the adolescent counterpart of both these adult syndromes). The only exceptions to the findings of predictive specificity over 10 years were for the males' scores on the Thought Problems and Attention Problems syndromes on the YASR, which were most strongly predicted by the YSR Anxious/Depressed and Thought Problems syndromes, respectively. The YASR Internalizing, Externalizing, and Total Problems scores were also most strongly predicted by their YSR counterparts. An additional finding was that significantly more participants whose YSR Total Problems scores were in the clinical range met criteria for DSM-IV diagnoses in structured interviews 10 years later than participants whose YSR Total Problems scores were not in the clinical range. Several YSR syndrome scales also predicted particular kinds of DSM-IV diagnoses over the 10-year period.

Prediction of Adult Syndrome Scores from Child/Adolescent ASEBA Scores in Clinical Samples

American Clinical Sample: 8-Year Predictive Relations. The foregoing studies supported the construct validity of the adult scales in terms of predictive relations from adolescent YSR and CBCL scores to adult YASR and YABCL scores over periods of 3 to 10 years in American and Dutch samples that were randomly selected to be representative of their respective populations. The construct validity of the adult syndromes has also been supported by predictive relations found for adults who had received mental health services when they were children or adolescents (Stanger, MacDonald, McConaughy, & Achenbach, 1996). Structural equation modeling of the latent variables jointly measured by YASR and YABCL syndrome scores showed that CBCL syndrome scores obtained at referral 8 years earlier were the most specific predictors of their analogous adult syndrome constructs. In addition, the adult Intrusive syndrome (designated as *Shows Off* in the Stanger et al. article) was strongly predicted by the CBCL Aggressive Behavior syndrome.

The strong predictive relations between pre-adult scores on the Aggressive Behavior syndrome and adult scores on the Intrusive syndrome in the clinical follow-up sample are consistent with findings in the American national sample (Achenbach et al., 1995c). Specifically, in longitudinal path analyses, the strongest predictors of the mean of adult Intrusive syndrome scores for 707 participants were adolescent Aggressive Behavior syndrome scores, averaged across CBCL and YSR ratings. This does not meant that the Intrusive syndrome was merely the adult version of the child/adolescent Aggressive Behavior syndrome, because the child/adolescent Aggressive Behavior syndrome was also the strongest predictor of the adult Aggressive Behavior syndrome.

The findings thus indicate that the child/adolescent Aggressive Behavior syndrome becomes differentiated into two syndromes on the adult instruments: One adult syndrome, the Aggressive Behavior syndrome, comprises overtly aggressive behaviors; the second adult syndrome, designated as Intrusive, comprises behaviors that may be socially obnoxious but are not overtly aggressive. According to these findings, during the transition to adulthood, some aggressive adolescents may retain the intrusive behaviors of the child/adolescent Aggressive Behavior syndrome but manifest less overt aggressive behavior. Other aggressive adolescents, however, may continue to manifest the overtly aggressive behavior that comprises the adult Aggressive Behavior syndrome.

Dutch Clinical Sample: 10.5-Year Predictive Relations. A similar study assessed Dutch adults who had received mental health services at a mean of 10.5 years earlier (Visser, van der Ende, Koot, & Verhulst, 2000). In this study, scores for each YASR syndrome were regressed separately on initial scores for all eight CBCL syndromes and for all eight YSR syndromes. Scores for each YABCL syndrome were also regressed on each score for all eight CBCL syndromes. There were thus 24 separate regressions of adult syndrome scores on childhood scores obtained at intake into mental health services (8 syndromes x 3 combinations of respondents). Age, gender, SES, and length of follow-up interval were also included as candidate predictors in the regression analyses. In 22 of the 24 regression analyses, the adult syndromes were most strongly predicted by their counterpart syndromes on

the CBCL and YSR (with the childhood Aggressive Behavior syndrome being the counterpart of the adult Intrusive syndrome). The two exceptions were the YASR Somatic Complaints syndrome and the YASR Thought Problems syndrome. These syndromes were most strongly predicted by the YSR Anxious/Depressed and Social Problems syndromes, respectively.

Relations of Adult Scales to Diagnostic Assessment

American Clinical Sample. In a second follow-up of 181 of the Stanger et al. (1996) young adults who had received mental health services as children or adolescents, DSM-III-R diagnoses were made on the basis of structured diagnostic interviews (Achenbach, McConaughy, LaRiviere, & Stanger, 1997). The interviewers also completed the *Global Assessment of Functioning* scale (GAF; American Psychiatric Association, 1987) on the basis of the diagnostic interviews and life history interviews with the participants. The participants completed YASRs either prior to the interviews or after the interviews, in counterbalanced order. Interviewers did not know the participants' YASR scores.

The YASR Total Problems score correlated .67 with the number of DSM diagnoses and -.65 with GAF scores (negative *r* because high GAF scores indicated good functioning, whereas high YASR problem scores indicated poor functioning). There were also many significant point-biserial *r*s between YASR syndrome scales and DSM diagnoses, which were scored as present vs. absent. The highest point-biserial *r*s were .60, .47, and .46 between the YASR Anxious/Depressed syndrome and diagnoses of Generalized Anxiety Disorder, Post-Traumatic Stress Disorder, and Agoraphobia/Social Phobia, respectively; a point-biserial *r* of .53 between the YASR Aggressive Behavior syndrome and Oppositional Defiant Disorder (which was included in the DSM diagnostic interview, even though it is often attributed only to pre-adults); and a point-biserial *r* of .47 between the YASR Delinquent Behavior syndrome and Antisocial Personality Disorder. Although they were empirically derived from the "bottom up" by factor analysis, the YASR syndromes were thus significantly associated with certain DSM diag-

nostic constructs, which were formulated using a "top-down" approach.

American Community Sample. In an American community sample that had been studied longitudinally since age 5, the Diagnostic Interview Schedule (DIS) was administered at age 18 years (Giaconia, Reinherz, Paradis, Hauf, & Stashwick, 2001). At age 21, the participants completed the YASR and their mothers completed the YABCL. Analyses of age 21 outcomes for diagnoses that had been made at age 18 showed that participants who had been diagnosed as having Major Depressive Disorder at age 18 obtained significantly higher Internalizing scores on both the YASR and YABCL at age 21 than participants who had not received this diagnosis. Participants who had been diagnosed as having either Major Depressive Disorder or Drug Abuse/Dependence Disorder at age 18 obtained significantly higher YASR Externalizing scores at age 21 than participants who did not receive these diagnoses.

In the same sample, YASR Internalizing, Externalizing, and Total Problems scores were significantly higher for women who had been sexually abused before age 18 than for women who had not reported sexual abuse (Silverman, Reinherz, & Giaconia, 1996). Some of these YASR scales were also significantly higher for men and women who had been physically abused than for those who had not reported physical abuse.

American Abused Sample. In a sample of adults who reported being physically or sexually abused in childhood, Posttraumatic Stress Disorder (PTSD) symptoms correlated .86 with YASR Total Problems scores and .80 with YASR Internalizing scores (Muller, Lemieux, & Sicoli, 2001). In addition, the Beck Depression Inventory (BDI) correlated .78 with YASR Total Problems and .75 with YASR Internalizing. (Laumann-Billings and Emery, 2002, have reported *r* = .74 between the BDI and Internalizing scored from the YASR and YSR.) The Beck Anxiety Inventory correlated .69 and .63 with YASR Total Problems and Internalizing, respectively. Negative views of the self scored from the Relationship Scales Questionnaire correlated higher (*r* = .61) with YASR Internalizing scores than with either of the Beck scales or with PTSD symptoms.

Dutch General Population Sample. In a Dutch sample that had been randomly selected, adults who completed the YASR were interviewed later with the *Schedules for Clinical Assessment in Neuropsychiatry* (SCAN; Ferdinand et al., 1995), covering 61 DSM-III-R diagnoses. The YASR Total Problems scores yielded a mean *r* of .74 with total symptom scores on the SCAN, averaged across men and women. Furthermore, participants who met criteria for DSM-III-R diagnoses on the SCAN obtained significantly higher Total Problems, Internalizing, and Externalizing scores on the YASR than did participants who did not meet criteria for diagnoses. In addition, the YASR Total Problems score yielded an *r* of -.62 with the GAF, which is similar to the *r* of -.65 found in the follow-up of American adults who had previously received mental health services (Achenbach et al., 1997). Thus, the associations of YASR Total Problems with diagnostic interviews and with GAF ratings were similar in an American clinical sample and a Dutch general population sample.

When reassessed 6 years later, far more of the Dutch adults who obtained deviant YASR Total Problems scores (combined borderline and clinical ranges) met criteria for interview-based DSM-IV diagnoses than adults whose YASR scores were in the normal range (52% vs. 7%, *p*<.001; Hofstra, van der Ende, & Verhulst, 2002)

Discriminative Performance of the Adult Scales vs. the GHQ and SCL-90-R among Dutch Adults

Two Dutch studies have compared the discriminative power of the YASR with two widely used measures of adult psychopathology, the *General Health Questionnaire* (GHQ; Goldberg, 1992) and the *SCL-90-R* (Derogatis, 1977). In one study, all three instruments were completed by 594 adults who had not received mental health services in the preceding year and by 294 who were referred for mental health services (Wiznitzer et al., 1992). Receiver Operating Characteristic analyses (ROC; Swets & Pickett, 1982) showed that the 90th percentile cutpoint for the clinical range on the YASR Total Problems score discriminated between referred vs. nonreferred adults signifi-

cantly better than the GHQ and nonsignificantly better than the SCL-90-R. The findings were similar for both inpatients and outpatients. As a measure of general psychopathology, the construct validity of the YASR Total Problems was thus supported by its ability to discriminate between referred and nonreferred Dutch adults significantly better than one widely used measure, the GHQ, and nonsignificantly better than another, the SCL-90-R. (The YASR Total Problems score correlates .99 with the ASR Total Problems score.)

The second Dutch study compared the ability of the YASR, GHQ, and SCL-90-R to predict multiple criteria of maladjustment across 2 years in a randomly selected general population sample assessed at ages 18 to 22 (Ferdinand & Verhulst, 1994). The total problems scores on all three instruments significantly predicted referral for mental health services and the participants' reports of needing help that they had not actually obtained. Internalizing, Externalizing, and some syndrome scales of the YASR also predicted police contacts, alcohol abuse, and suicidal behavior and ideation. Although the GHQ and SCL-90-R also predicted some of these signs of disturbance, neither of them predicted all the signs that were predicted by the YASR.

Correlations of YASR Scales with MMPI-2 Scales among Turkish Adults

The original MMPI scales were constructed partly by selecting items that discriminated between psychiatric patients diagnosed according to diagnostic categories of the 1930s. The MMPI-2 incorporates modernized versions of some of the original items, plus renormed scales (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989). The following features make the MMPI-2 very different from the YASR: The MMPI-2 scales are based on 1930s diagnostic categories; the 567 MMPI-2 items tap personality and attitudinal characteristics that were chosen by an empirical keying method mainly for their associations with diagnoses rather than for their value in describing behavioral/ emotional problems that are potentially observable by others; and the MMPI-2 items are scored dichotomously as true vs. false.

Because they assess functioning in potentially complementary ways, the ASEBA adult forms and the MMPI-2 may be useful to administer together in clinical and research applications. To test associations between ASEBA and MMPI-2 scales, correlations were computed between Turkish translations of both instruments. The vicissitudes of translating such different instruments from idiomatic American English to a language as different as Turkish provided a strong test of whether there are significant associations between scales of the two instruments.

The Turkish YASR and MMPI-2 were completed by 70 19- to 28-year-old students from several Turkish universities (unpublished data provided by Drs. Nese Erol and Isik Savasir). Table 10-9 displays *r*s between YASR and MMPI-2 scales that were significant at *p*<.01. To flag correlations that might have been significant by chance in the 99 analyses (11 YASR scales x 9 MMPI-2 scales), superscript *a* marks the four smallest *r*s that were significant at *p*<.01. (Four out of 99 analyses are expected to be significant by chance at *p*<.01, with a *p* = .01 protection level; Sakoda et al., 1954).

Despite the differences between the YASR and MMPI-2, plus the effects of translation, Table 10-9 reveals several associations that were large according to Cohen's (1988) criteria, i.e., *r*s ≥.50: YASR Anxious/Depressed with MMPI-2 Depression, Paranoia, Psychasthenia, and Schizophrenia; YASR Attention Problems with MMPI-2 Psychasthenia; YASR Aggressive Behavior with MMPI-2 Paranoia, Psychasthenia, and Schizophrenia; YASR Internalizing with MMPI-2 Paranoia, Psychasthenia, and Schizophrenia; YASR Externalizing with MMPI-2 Schizophrenia; and YASR Total Problems with MMPI-2 Paranoia, Psychasthenia, and Schizophrenia.

The MMPI-2 Psychasthenia and Schizophrenia scales had especially numerous large correlations with YASR scales. The correlations ranged up to .73 with YASR Total Problems, suggesting that these MMPI-2 scales reflect a general dimension of psychopathology like the YASR Total Problems score. Findings that reflect greater specificity in relations between YASR and MMPI-2 scales include the significant *r*s of YASR Anxious/Depressed with MMPI-2 Depression and

Paranoia; YASR Withdrawn with MMPI-2 Social Introversion; YASR Somatic Complaints with MMPI-2 Hypochondriasis; and YASR Aggressive Behavior and Rule-Breaking (Delinquent) Behavior with MMPI-2 Psychopathic Deviate.

Scores for Finnish Adults Whose Families Received Counseling 15-20 years Earlier and Who were Assessed for Depression 10 Years Earlier

Families of Finnish children randomly selected from a birth cohort were assigned to receive 5 years of family counseling or to a no-treatment control condition (Aronen & Arajarvi, 2000; Aronen & Soininen, 2000). At age 11, the children completed the Children's Depression Inventory (CDI). At ages 20 to 21, they completed the YASR and the Beck Depression Inventory (BDI).

High age 11 scores on the CDI significantly predicted high adult scores on the YASR Total Problems, Anxious/Depressed, and Aggressive Behavior scales. Low age 11 CDI scores, by contrast, significantly predicted favorable scores on the YASR Friends, Job, Education, and Mean Adaptive scales. However, even though both the CDI and BDI are intended to assess depression, the CDI did not significantly predict BDI scores.

The long-term effects of the counseling intervention were evaluated by comparing the age 20-21 YASR and BDI scores obtained by adults whose families had received either the counseling or the control condition during the first 5 years of the adults' lives. It was found that the adults whose families had received counseling obtained significantly lower scores on the YASR Total Problems, Internalizing, Anxious/Depressed, and Attention Problems scales and on the BDI. In addition, adults whose families had been rated as high-risk but who had received counseling obtained significantly lower YASR Total Problems scores than adults from high-risk families in the control condition. For low-risk and high-risk groups combined, the proportion of adults obtaining YASR Total Problems scores in the borderline and clinical ranges was significantly smaller in the counseling group than the control group. The YASR was thus sensitive to differences in

Table 10-9
Significant (*p*<.01) Correlations of YASR Scales with MMPI-2 Scales

YASR Scales	Hypochondriasis	Depression	Hysteria	MMPI Scales Psychopathic	Paranoia	Psychasthenia	Schizophrenia	Hypomania	Social Introversion
Anxious/Depressed	.35	.50	—	.33	.50	.69	.64	—	.42
Withdrawn	—	.31[a]	—	—	.38	.41	.39	—	.43
Somatic Complaints	.43	—	—	—	.39	.39	.36	—	—
Thought Problems	—	—	—	—	—	—	—	—	—
Attention Problems	.37	.38	—	—	.37	.53	.47	—	—
Aggressive Behavior	.32[a]	—	—	.44	.44	.47	.54	—	—
Rule-Breaking Behavior[b]	—	—	—	.43	.33[a]	.33	.46	.38	—
Intrusive	—	—	—	—	—	—	—	—	—
Internalizing	.34	.49	—	.33[a]	.51	.67	.62	—	.47
Externalizing	—	—	—	.43	.41	.43	.53	.38	—
Total Problems	.42	.45	—	.47	.61	.73	.71	—	.38

Note. Participants were 70 Turkish university students who completed Turkish translations of the YASR and MMPI-2. The authors are grateful to Drs. Nese Erol and Isik Savasir for use of their data.

[a]Not significant when corrected for the number of analyses (Sakoda et al., 1954).

[b]Designated as Delinquent Behavior on 1997 YASR.

functioning that apparently resulted from family counseling 15 to 20 years earlier.

SUMMARY

This chapter presented a variety of evidence for the validity of ASEBA adult scores. The *content validity* of the problem items was supported by *(a)* their ability to discriminate significantly between referred and nonreferred samples, and/or *(b)* their significant loadings on empirically based syndromes, and/or *(c)* their identification by experts as being very consistent with DSM-IV diagnostic categories. Nearly all of the adaptive functioning items also discriminated significantly between referred and nonreferred samples.

Criterion-related validity of the adult scale scores was supported by the significant association of all but the ABCL Intrusive syndrome scores with referral status independently of demographic effects. Categorical analyses via odds ratios and chi squares showed that classification of scores in the normal vs. combined borderline and clinical ranges was significantly associated with referral status for nearly all ASR and ABCL scales. Discriminant analyses showed that the best cross-validated classification rate (87% correctly clas-

sified as referred vs. nonreferred) was achieved by the ASR problem items as candidate predictors. Item *103. Unhappy, sad, or depressed* was the priority predictor of referral status in analyses of both the ASR and ABCL, as it has been in many analyses of preadult samples. As shown in Table 10-7, the probability that particular Total Problems scores were from referred samples can be used to estimate the likelihood that an individual's Total Problems score is high enough to be of concern.

The complex issue of *construct validity* was addressed with a variety of evidence for associations with other measures of psychopathology reported in studies from Finland, the Netherlands, Turkey, and the United States. The evidence included significant predictive relations from ASEBA child and adolescent syndrome scores to the corresponding syndromes assessed with the ASEBA adult instruments over periods up to 10.5 years in general population and clinical samples; significant associations between ASEBA adult scales and diagnostic assessment; significant associations with the Beck Depression Inventory, the Beck Anxiety Inventory, the MMPI, and the SCL-90-R; and significant associations with an intervention 15 to 20 years earlier and with Child Depression Inventory scores obtained at age 11.