Psychology in the Schools, Vol. 38(3), 2001 © 2001 John Wiley & Sons, Inc.

STABILITY OF THE ADJUSTMENT SCALES FOR CHILDREN AND ADOLESCENTS

GARY L. CANIVEZ, AMANDA R. PERRY, AND ELIZABETH M. WELLER

Eastern Illinois University

Investigation of the short-term (90-day) stability of the Adjustment Scales for Children and Adolescents is reported for 124 randomly selected children in grades ranging from kindergarten to grade 12. Significant test-retest stability coefficients were obtained and mean differences across the retest interval did not exceed .8 raw score points. The Solitary Aggressive-Impulsive, Diffident, and Lethargic/Hypoactive syndromes and the global Underactivity scale showed significant raw score and T score changes across the retest interval, but the effect strengths were small. Syndromic Profile Classifications and Discriminant Classifications were also significantly consistent across the retest interval. Results were similar to those obtained in other stability studies of teacher report behavior rating scales. © 2001 John Wiley & Sons, Inc.

School psychologists prefer objective assessment methods that facilitate links between assessment and intervention (Reschly & Ysseldyke, 1995), rather than inferential methods, in assessing psychopathology and problem behavior. Consequently, standardized behavior rating scales and checklists have achieved great popularity among school and clinical psychologists (Merrell, 1994a).

Among school psychologists, behavior rating scales are the most frequently used instruments in assessing emotional and behavioral difficulties of youths (Stinnett, Havey, & Oehler-Stinnett, 1994). Behavior rating scales are "one of the most efficient, sound, and effective ways . . . to identify a referred student's behavioral strengths and weaknesses . . ." (Knoff, 1995, p. 857) and their use has also been designated a "best practice" in the assessment of emotional and behavioral disorders (McConaughy & Ritter, 1995).

In addition to their more objective method, behavior rating scales allow relatively unobtrusive evaluations of student behaviors in natural social settings such as schools, classrooms, and homes. Within the classroom and other school settings, teachers are natural observers and informants since they have the comparative experience of observing many students across time and varied social contexts. As such, they seem to take a normative perspective in rating difficulties in children. Consequently, teachers have often been considered among the most accurate adult raters of child behavior (Kamphaus & Frick, 1996). Behavior rating scales are also cost-effective methods for assessing different behaviors across different environments and raters.

Although behavior rating scales have many positive qualities, there are potential threats to their validity such as rater bias (i.e., halo effect, leniency error), rater competency, relevant contact, and rater agreement. Rater agreement may not necessarily be a problem as it is possible that behaviors may vary with respect to different environments and the scale may simply measure those differences. School psychologists are generally aware of classroom differences and the impact on behaviors but need to keep this issue in mind in the assessment process. Many behavior rating scales have also been limited by poor standardization samples that lack national representation.

The Adjustment Scales for Children and Adolescents (ASCA; McDermott, Marston, & Stott, 1993) is a relatively new behavior rating scale system designed to assess youth psychopathology based on teacher reporting of child behaviors within school settings. The ASCA has a nationally representative standardization sample, and evidence of acceptable score reliability and validity is presented in the *ASCA Manual* and other publications (Canivez, in press). The ASCA contains 156 items, 97 that are scorable for dimensions of psychopathology and, based on factor analyses, are

Correspondence to: Gary L. Canivez, PhD, Department of Psychology, 600 Lincoln Avenue, Charleston, IL 61920-3099. E-mail: cfglc@eiu.edu; Internet: http://www.ux1.eiu.edu/~cfglc.

singularly assigned to one of six core syndromes or two supplementary syndromes. The six core syndromes, which have been found to be reliable across gender, age, and race/ethnicity, include Attention Deficit/Hyperactive (ADH), Solitary Aggressive-Provocative (SAP), Solitary Aggressive-Impulsive (SAI), Oppositional-Defiant (OPD), Diffident (DIF), and Avoidant (AVO). These six core syndromes also combine to form two composite (second-order) or overall adjustment indexes: Overactivity (ADH, SAP, SAI, and OPD syndromes) and Underactivity (DIF and AVO syndromes). Delinquency (DEL) and Lethargic-Hypoactive (LEH) make up the two supplementary syndromes that are reliable for certain subgroups in the population. Core syndromes, supplementary syndromes, and overall adjustment scales are reported as normalized T scores (M = 50, SD = 10) and percentiles.

McDermott (1994) presents extensive reliability and validity evidence in the ASCA Manual. Internal consistency estimates for the total standardization sample ranged from .68 to .86 for the six core syndromes and two supplementary syndromes. Alpha coefficients equaled .92 for the Overactivity scale and .82 for the Underactivity scale. Exploratory and confirmatory analyses support the factor structure of the ASCA at the item, core syndrome, and second-order levels. Convergent and divergent validity studies comparing the ASCA with the Conners Teacher Rating Scale (CTRS; Trites, Blouin, & Laprade, 1982) and the Child Behavior Checklist (CBCL; Achenbach & Edelbrock, 1983) found significant correlations among similar psychological dimensions (McDermott, 1994). In general, psychometric characteristics of the ASCA are acceptable and meet standards for both group and individual decision making (Salvia & Ysseldyke, 1995).

Evidence of short-term (30-school-day) stability reported in the ASCA Manual was based upon a sample of 40, 14 to 17-year-old female students in Pennsylvania (McDermott, 1994). All correlations were significant and there were no significant differences in T scores across the retest interval. Test-retest reliabilities ranged from .66 to .91 for the six core syndromes (n = 40) and equaled .75 for the Overactivity scale and .79 for the Underactivity scale. In another small study of ASCA short-term (45-day) stability (n = 51), Canivez (2000) also found significant stability coefficients; however, several syndromes showed significant yet small changes across the retest interval.

Previous investigations of ASCA short-term stability pertained to the obtained *T* scores and their resulting cut score interpretations. In addition to cut score interpretations, McDermott (1994) also presented two multivariate methods of interpretation: *Syndromic Profile Classification* and *Discriminant Classification*. The Syndromic Profile Classification approach allows clinicians the ability to consider all six core syndromes simultaneously to aid in differential diagnosis. Discriminant Classification allows clinicians the ability to determine whether or not the profile better resembles a normal student or a student classification methods has yet to appear in the empirical literature.

Syndromic Profile Classification is based on results of the cluster analysis of the standardization sample that produced 22 profile types (McDermott, 1994; McDermott & Weiss, 1993, 1995). Syndromic Profile Classification involves comparing a youth's core syndrome T scores to the mean core syndrome T scores for one or more of the 22 profile types (14 major types and 8 clinical subtypes) to determine which profile is most similar. Classification of the youth's profile is based on the generalized distance score (GDS) method (McDermott, 1994). Canivez (1996, 1998a) automated the calculations for the GDS to assure reliable computation.

Discriminant Classification is based on results of discriminant function analyses conducted in which the ASCA was found to correctly classify normal from socially/emotionally disturbed youths (McDermott, 1994; McDermott et al., 1995), and which yielded positive predictive power estimates exceeding a recommended standard (.75) for diagnostic tests (Milich, Widiger, & Landau, 1987). Discriminant Classification involves applying the youth's six core syndrome T scores to

Stability of the ASCA

the two linear discriminant function regression equations that are used to determine which group (normal vs. socially/emotionally disturbed) the youth in question is most likely to belong. The youth's profile is classified as most similar to the group that results in the higher discriminant score. Canivez (1996, 1998a) also provides automated calculation for Discriminant Classifications.

Given the potential diagnostic applications of the ASCA, independent assessment of ASCA stability with a larger sample is needed. The purpose of this study was to examine the short-term test-retest stability of the ASCA with a larger sample of students enrolled in regular education programs. Additionally, the present study sought to replicate and extend previous results by also examining the stability of Syndromic Profile Classifications and Discriminant Classifications in addition to stability of ASCA raw scores and *T* scores.

METHOD

Participants

Thirteen public school teachers in two rural Illinois public school districts volunteered to participate in the present study after being presented with details of the procedure. All teachers were female and certified by the State Department of Education. Students' grade levels ranged from kindergarten to grade 12, but there were no students in grades seven or eight. Teachers agreed to rate twice 10 randomly selected children (5 male, 5 female) on the ASCA. Of the 13 teachers, 10 twice rated all 10 children while three teachers twice rated only eight children each. Thus, the total sample included 124 children (67 male, 57 female) ranging in age from 5 to 19 years (M = 10.77, SD = 3.60). Of these 124 children, 35 did not have race/ethnicity data provided by their teacher. For those children whose race/ethnicity was indicated, 71 (79.8%) were Caucasian American, 2 (2.2%) were Black/African American, 12 (13.5%) were Hispanic/Latino American, 2 (2.2%) were Native American, 1 (1.1%) was Asian American, and 1 (1.1%) was Bosnian. None of the students were reportedly classified as disabled.

Instrument

The Adjustment Scales for Children and Adolescents (ASCA; McDermott, Marston, & Stott, 1993) is a major revision of the Bristol Social Adjustment Guides (Stott, 1966; Stott, Marston, & Neill, 1975), for which there are British and Canadian norms. Psychopathology is conceived of and defined as multisituational expression of problem behaviors, as opposed to being defined by general or global ratings of problem behavior. The ASCA is appropriate for all noninstitutionalized youths ages 5 through 17 and contains 156 behavioral descriptions within 29 specific situations where teachers may observe students' behaviors. Specific behaviors (i.e., accepts the rules, plays for himself only, inclined to cheat, etc.) are listed within a specific situation (i.e., Does he play fairly?) to provide a context for the teacher's ratings. Thus, teachers do not attempt to judge the general frequency or intensity of a symptom as is typically done in rating scales such as the Behavior Assessment System for Children (BASC; Reynolds & Kamphaus, 1992), the Preschool and Kindergarten Behavior Scales (PKBS; Merrell, 1994b), or the Devereux Behavior Rating Scale–School Form (DBRS-SF; Naglieri, LeBuffe, & Pfeiffer, 1993).

The ASCA consists of six core syndromes (ADH, SAP, SAI, OPD, DIF, and AVO) and two supplementary syndromes (DEL and LEH). The core syndromes are combined to form two composite indexes (broad-band/global scales): Overactivity (ADH, SAP, SAI, and OPD syndromes) and Underactivity (DIF and AVO syndromes). These global scales are similar to the externalizing (or conduct problem) and internalizing (or withdrawal) dimensions consistently found in the majority of child psychopathology measures and in the literature on childhood psychopathology (Quay, 1986).

Canivez et al.

Raw scores are converted to normalized T scores based on the nationally representative standardization sample. The ASCA was normed on a representative national sample of 1,400 youths, blocked according to gender, age, and grade level, and stratified proportionately according to national region, community size, race/ethnicity, parent education, family structure, and handicapping condition. The ASCA was also co-normed with the Differential Abilities Scale (Elliot, 1990) by The Psychological Corporation.

Procedure

The 13 participating teachers were instructed how to randomly select and then rate 10 students (5 male, 5 female) attending their class on the ASCA following the standard administration procedures. The teachers again rated the students 90 days later. As students were randomly selected from regular classrooms it was thought that a 90-day retest interval would provide a more stringent (yet reasonable) test of stability than previously used retest intervals. All ASCA rating forms were returned to the second and third authors who scored them according to standard procedures (McDermott, 1994). Core Syndrome, Supplementary Syndrome, and global Adjustment Scale *T* scores were obtained from the *ASCA Manual*.

Syndromic Profile Classifications were made using the generalized distance score (GDS) method according to the ASCA Manual (McDermott, 1994) utilizing an automated scoring template (Canivez, 1996, 1998a). The GDS provides a measure of profile similarity (dissimilarity) by examining deviations of a youth's core syndrome T scores from the average T scores for a specified group (ASCA profile type). The youth's profile is classified as most similar to the ASCA profile type that results in the smallest GDS. Broad classifications based on syndromic profile types suggested in the ASCA Manual were also made and examined for stability. As indicated in the ASCA Manual (pp. 24–25) and based on T score elevations, profile Type 1 is classified as Adjusted, Types 2 through 5 are classified as Adequately Adjusted, Types 6 through 12 are classified as Marginally Adjusted, Types 13 through 18 are classified as At Risk, and Types 19 through 22 are classified as Maladjusted. Stability among these five classification categories was assessed. Additional reductions of these five categories were performed to investigate further effects on stability. Adjusted, Adequately Adjusted, and Marginally Adjusted groups were combined into an Adjusted category and stability compared with the At Risk and Maladjusted groups. Finally, the At Risk and Maladjusted groups were combined into a classification termed Not Adjusted and stability for Adjusted and Not Adjusted groups were examined. Table 1 presents the classifications of each of the 22 syndromic profiles into the 5, 3, and 2 category groupings.

Discriminant Classifications were also made according to the ASCA Manual using linear discriminant classification equations (McDermott, 1994, p. 29) provided in an automated scoring template (Canivez, 1996, 1998a). Profiles were classified normal or socially/emotionally disturbed based on the regression equation resulting in the highest discriminant score. Stability for Discriminant Classifications was also investigated.

Data Analyses

Pearson product-moment correlation coefficients between first and second ratings were calculated for raw scores and *T* scores obtained for the ASCA Core Syndromes, Supplementary Syndromes, and overall Adjustment Scales to assess the *direction* of stability (McDermott, 1988). Dependent *t* tests were conducted to investigate changes in ratings across the 90-day retest interval to assess the *level* of stability (McDermott, 1988). Effect strengths of mean rating changes across the retest interval were estimated using η^2 , an index of the proportion of variability explained by the effect across the retest interval (Kiess, 1996) to assess the clinical significance of changes in ratings.

Table 1Broad Classifications of the 22 Syndromic Profiles into	5, 3, and 2 Broad Classification Le	vels	
ASCA Syndromic Profile Type	5-Level Classification	3-Level Classification	2-Level Classification
1. Good Adjustment	Adjusted	Adjusted	Adjusted
2. Adequate Adjustment w/ Inhibition	Adequate Adjustment	Adjusted	Adjusted
3. Adequate Adjustment w/ Disruptiveness	Adequate Adjustment	Adjusted	Adjusted
4. Adequate Adjustment w/ Apprehension	Adequate Adjustment	Adjusted	Adjusted
5. Adequate Adjustment w/ Indifference	Adequate Adjustment	Adjusted	Adjusted
6. Adequate Adjustment w/ Withdrawal	Marginal Adjustment	Adjusted	Adjusted
7. Adequate Adjustment w/ Motivation Deficit	Marginal Adjustment	Adjusted	Adjusted
8. Adequate Adjustment w/ Avoidance	Marginal Adjustment	Adjusted	Adjusted
9. Adequate Adjustment w/ Attention Seeking	Marginal Adjustment	Adjusted	Adjusted
10. Adequate Adjustment w/ Moodiness	Marginal Adjustment	Adjusted	Adjusted
11. Adequate Adjustment w/ Nonparticipation	Marginal Adjustment	Adjusted	Adjusted
12. Adequate Adjustment w/ Dependency	Marginal Adjustment	Adjusted	Adjusted
13. Undersocialized Aggressive	At Risk	At Risk	Not Adjusted
14. Oppositional	At Risk	At Risk	Not Adjusted
15. Provocative, Attention Seeking	At Risk	At Risk	Not Adjusted
16. Provocative, Manipulative	At Risk	At Risk	Not Adjusted
17. Impulsive Aggressive	At Risk	At Risk	Not Adjusted
18. Attention-Deficit Hyperactive	At Risk	At Risk	Not Adjusted
19. Instrumental Aggressive	Maladjusted	Maladjusted	Not Adjusted
20. Defiant Aggressive	Maladjusted	Maladjusted	Not Adjusted
21. Avoidant	Maladjusted	Maladjusted	Not Adjusted
22. Schizoid with Depressed Mood	Maladjusted	Maladjusted	Not Adjusted

Canivez et al.

Stability of Syndromic Profile Classifications and Discriminant Classifications were examined through the use of kappa coefficients (Canivez, 1998b; Cohen, 1960) and z tests due to the nominal scale classifications that result from these two methods. Unlike the core syndrome, supplementary syndrome, and global adjustment scale T scores, Syndromic Profile Classifications and Discriminant Classifications are nominal scale variables. When investigating agreement on nominal scale or categorical variables, nominal scale statistics such as kappa (κ) should be utilized (Cohen, 1960; Fleiss, 1981; McDermott, 1988). Kappa provides an index of agreement beyond chance and is interpreted much like a correlation coefficient as it ranges from -1 to +1.

RESULTS

Tables 2 and 3 present the stability coefficients, descriptive statistics, t tests, and retest interval effect strengths for ASCA raw scores and T scores, respectively. All stability coefficients were statistically significant (p < .0001). Stability coefficients ranged from .51 to .78 (Mdn = .69) for raw scores and ranged from .48 to .68 (Mdn = .61) for T scores. Most syndrome raw scores and T scores showed no significant mean changes across the retest interval. For raw scores, the SAI, DIF, and LEH syndromes and the Overactivity and Underactivity adjustment scales showed significant changes across the retest interval (see Table 2). Effect strengths, however, were small and mean differences were less than .8 raw score points. For T scores, the SAI, DIF, and LEH syndromes and the Underactivity adjustment scale showed significant changes across the retest interval (see Table 3). However, as with raw scores, the effect strengths were small. Figure 1 illustrates the mean ASCA T score profiles across the retest interval.

Stability of the Syndromic Profile Classifications and their broad classifications is summarized in Table 4. All kappa coefficients were significant, indicating that classifications of profiles

Retest Interval Effect Stre	ngths for	ASCA K	Raw Score	25					
	n	r	First Testing		Second Testing				
			М	SD	М	SD	t	р	η^2
Core syndromes									
ADH	124	.74	2.51	3.81	2.15	3.34	1.56	.123	.02
SA(P)	124	.78	0.54	1.38	0.48	0.99	0.82	.411	.01
SA(I)	124	.59	0.21	0.75	0.09	0.38	2.22	.028	.04
OPD	124	.65	0.89	1.95	0.65	1.33	1.76	.081	.02
DIF	124	.75	1.75	2.34	1.30	2.02	3.17	.002	.08
AVO	124	.51	1.05	1.50	0.91	1.44	1.05	.295	.01
Supplementary syndromes									
DEL	86	.53	0.24	0.70	0.14	0.41	1.63	.106	.03
LEH	80	.64	0.88	1.69	0.46	0.94	2.82	.006	.09
Adjustment scales									
OVR	124	.77	4.15	6.76	3.36	4.96	2.01	.047	.03
UNR	124	.74	2.80	3.06	2.21	2.66	3.13	.002	.07

 Table 2

 Stability Coefficients, Descriptive Statistics, t tests, and

 Retest Interval Effect Strengths for ASCA Raw Scores

Note. ADH = Attention Deficit-Hyperactive, SA (P) = Solitary Aggressive (Provocative), SA (I) = Solitary Aggressive (Impulsive), OPD = Oppositional Defiant, DIF = Diffident, AVO = Avoidant, DEL = Delinquent, LEH = Lethargic (Hypoactive), OVR = Overactivity, UNR = Underactivity. All correlations significant, p < .0001.

222

	n	r	First Testing		Second Testing				
			М	SD	М	SD	t	р	η^2
Core syndromes									
ADH	124	.64	49.85	12.00	48.85	10.59	1.17	.246	.01
SA(P)	124	.65	50.20	9.99	50.58	9.94	0.51	.613	.00
SA(I)	124	.68	49.60	7.69	48.48	5.67	2.22	.029	.04
OPD	124	.49	49.86	12.18	48.98	9.15	0.88	.378	.01
DIF	124	.62	51.72	10.71	49.28	10.24	2.96	.004	.07
AVO	124	.48	51.15	10.23	49.75	10.07	1.51	.134	.02
Supplementary syndromes									
DEL	86	.50	48.62	9.07	47.85	7.91	0.83	.406	.01
LEH	80	.63	51.54	10.99	49.20	9.21	2.38	.020	.07
Adjustment scales									
OVR	124	.60	49.78	11.56	49.82	9.94	0.05	.963	.00
UNR	124	.57	51.87	10.47	49.38	9.66	2.96	.004	.07

 Table 3

 Stability Coefficients, Descriptive Statistics, t tests, and

 Retest Interval Effect Strengths for ASCA T Scores

Note. ADH = Attention Deficit-Hyperactive, SA (P) = Solitary Aggressive (Provocative), SA (I) = Solitary Aggressive (Impulsive), OPD = Oppositional Defiant, DIF = Diffident, AVO = Avoidant, DEL = Delinquent, LEH = Lethargic (Hypoactive), OVR = Overactivity, UNR = Underactivity. All correlations significant, p < .0001.

from Time 1 to Time 2 were generally stable when compared with chance levels of agreement. However, clinical significance for the stability of the 22 Syndromic Profiles was classified as poor (Cicchetti, 1994) or fair (Everitt & Hay, 1992; Landis & Koch, 1977). For the 5 Broad Classifications, agreement was classified as poor (Cicchetti, 1994) or moderate (Everitt & Hay, 1992; Landis & Koch, 1977). Clinical significance of agreement across the retest interval for the 3 and 2 Broad Classifications were classified as fair (Cicchetti, 1994) or moderate (Everitt & Hay, 1992; Landis & Koch, 1977).

Stability of the Discriminant Classifications is also summarized in Table 4. As with the Syndromic Profile Classifications, the Discriminant Classifications showed significant agreement from Time 1 to Time 2 ($\kappa = .35$, z = 3.93, p < .00008). Clinical significance, however, was considered poor (Cicchetti, 1994) or fair (Everitt & Hay, 1992; Landis & Koch, 1977). Of the 124 profiles, 91 (73%) were classified as "Normal" at Time 1 *and* at Time 2, while 10 (8%) were classified "SED" at Time 1 *and* at Time 2. Eleven profiles (9%) were classified "SED" at Time 1 and "Normal" at Time 2, while 12 (10%) were classified as "Normal" at Time 1 but "SED" at Time 2.

DISCUSSION

The present study investigated the 90-day stability of the Adjustment Scales for Children and Adolescents with a sample of students attending regular education classrooms and twice rated by their regular education classroom teacher. Test-retest stability coefficients across the 90-day interval were significant but somewhat lower in magnitude than those reported in the ASCA Manual with a 30-school-day retest interval (McDermott, 1994). This was expected due to the longer retest interval in the present study. The present study also found significant T score changes across the retest interval for the SAI, DIF, and LEH syndromes and the Underactivity adjustment scale,



FIGURE 1. Mean ASCA *T* score profiles for first and second testings. OVR = Overactivity, UNR = Underactivity, ADH = Attention Deficit-Hyperactive, SA(P) = Solitary Aggressive (Provocative), SA(I) = Solitary Aggressive (Impulsive), OPD = Oppositional Defiant, DIF = Diffident, AVO = Avoidant, DEL = Delinquent, LEH = Lethargic (Hypoactive).

but all showed small effect strengths and were considered clinically unimportant. McDermott (1994) found no significant changes across the retest interval for any of the syndromes or adjustment scales.

The test-retest stability coefficients in the present study, although significant, are slightly lower than those found for other teacher report child behavior rating scales across similar retest

	P.	P.	к	SE.	7	n
	10	1 0	ĸ	5.E _K	~	P
Syndromic Profile Classification						
22 Syndrome Profiles	.31	.08	.24	.03	9.56	.00001
5 Broad Classifications	.46	.23	.29	.05	6.19	.00001
3 Broad Classifications	.81	.62	.49	.07	6.82	.00001
2 Broad Classifications	.85	.64	.59	.09	6.63	.00001
Discriminant Classification						
Normal/SED (linear)	.81	.71	.35	.09	3.94	.00008

 Table 4

 Stability of ASCA Syndromic Profile Based Classifications and Discriminant Classifications

Note. P_o = proportion of observed agreement, P_c = proportion of chance agreement, κ = kappa, SE_{κ} = standard error of kappa. Copies of the Syndromic Profile Classification and Discriminant Classification analysis tables may be obtained from the first author.

Stability of the ASCA

intervals (Achenbach, 1991; Merrell, 1994b; Naglieri et al., 1993; Reynolds & Kamphaus, 1992). One possible reason for these differences may be found in how the items are scored. The ASCA items are endorsed and scored dichotomously (0-Absent, 1-Present), while other behavior rating scales like the Child Behavior Checklist 91–Teachers Report Form (CBCL-TRF; Achenbach, 1991), the Behavior Assessment System for Children–Teacher Rating Scales (BASC-TRS; Reynolds & Kamphaus, 1992), the Preschool and Kindergarten Behavior Scales (PKBS; Merrell, 1994b), and the Devereux Behavior Rating Scale–School Form (Naglieri et al., 1993) have items which are scored on a 3-, 4-, or 5-point continuum, thereby increasing variability at the item level as well as in the total scale or syndrome. Thus, one would expect higher correlations in these scales as a function of the greater available-item variability.

Another potential influence for the somewhat lower correlations among some scales is the number of items in the scale. As would be expected, lower correlations were observed in ASCA scales with fewer items.

This is the first study to systematically investigate the stability of the two multivariate interpretive classification methods presented in the *ASCA Manual*. Results indicated that the 22 Syndromic Profile Classifications and their resulting 5-, 3-, and 2-level broad classifications all demonstrated significant agreement across the retest interval indicating significant temporal stability. This is an encouraging and important finding to the extent that one would expect that the profile generated by a behavioral or psychopathology measure should be relatively stable over the short-term retest interval investigated in this study. Additionally, the Discriminant Classifications also showed significant temporal stability across the 90-day retest interval.

Although the Syndromic Profile Classifications and Discriminant Classifications showed significant stability, the clinical significance of the stability was fair or moderate only for the 3 and 2 Broad Classifications. A possible explanation for this is that many of the 22 Syndrome Profiles were not represented, and this can influence κ estimates. Increasing the sample size so that more or all of the 22 Syndrome Profiles would be represented would help provide a more adequate test of the stability of the 22 Syndrome Profiles.

Another significant contribution of this study is in the presentation of a method to investigate such nominal scale agreement in classification over time through the use of kappa and its significance test (Cohen, 1960; Fleiss, 1981). This same method has been applied in examining classification agreements between two independent raters (interrater reliability) on the ASCA (Canivez & Watkins, 2000).

Caution should be exercised in interpreting the results of this study as it is based on a small sample of students who are not representative of the population at large. Generalizability of these results is certainly limited, as the sample was predominantly Caucasian (79.8%) and included students from rural Illinois public schools. Another limitation is that the ratings were based on the reports of only 13 teachers who agreed to provide ratings across the 90-day retest interval. Future studies should continue to investigate the temporal stability of the ASCA in a similar manner as this study, but with larger and more diverse and representative student and teacher samples. Future studies should also continue to investigate reliability of Syndromic Profile Classifications and Discriminant Classifications, as these are potentially the most valuable diagnostic indicators on the ASCA.

References

Achenbach, T.M. (1991). Manual for the Teacher's Report Form and 1991 Profile. Burlington, VT: University of Vermont Department of Psychiatry.

Achenbach, T.M., & Edelbrock, C. (1983). Manual for the Child Behavior Checklist and Revised Child Behavior Profile. Burlington, VT: University of Vermont.

Canivez et al.

- Canivez, G.L. (in press). Review of the Adjustment Scales for Children and Adolescents. In J. Impara and B. Plake (Eds.), The fourteenth mental measurements yearbook. Lincoln, NE: Buros Institute of Mental Measurements, University of Nebraska.
- Canivez, G.L. (1996). Automated syndromic profile and discriminant classification analyses for the Adjustment Scales for Children and Adolescents (ASCA) v2.0. Microsoft[®] ExcelTM Spreadsheet Template for the Apple[®] MacintoshTM Microcomputer. Charleston, IL: Author.
- Canivez, G.L. (1998a). Automated syndromic profile and discriminant classification analyses for the Adjustment Scales for Children and Adolescents. Behavior Research Methods, Instruments, & Computers, 30, 732–734.
- Canivez, G.L. (1998b). Automated calculation of nominal scale agreement (Kappa) statistics for ASCA syndromic profile classifications and discriminant classifications. Microsoft® ExcelTM Spreadsheet Template for the Apple® MacintoshTM Microcomputer. Charleston, IL: Author.
- Canivez, G.L. (2000). Stability of regular education teacher ratings of normal and exceptional students on the Adjustment Scales for Children and Adolescents (ASCA). (ERIC Document Reproduction Service No. TM 030 644).
- Canivez, G.L., & Watkins, M.W. (2000). Interrater agreement for syndromic profile and discriminant classifications on the Adjustment Scales for Children and Adolescents (ASCA). Manuscript submitted for publication.
- Cicchetti, D.V. (1994). Guidelines, criteria, and rules of thumb for evaluating standardized assessment instruments in psychology. Psychological Assessment, 6, 284–290.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20, 37-46.
- Elliot, C.D. (1990). Differential Ability Scales: Introductory and technical handbook. San Antonio, TX: Psychological Corporation.
- Everitt, B.S., & Hay, D.F. (1992). Talking about statistics: A psychologist's guide to data analysis. New York: Wiley.
- Fleiss, J.L. (1981). Statistical methods for rates and proportions (2nd ed.). New York: Wiley.
- Kamphaus, R.W., & Frick, P.J. (1996). Clinical assessment of child and adolescent personality and behavior. Boston: Allyn and Bacon.
- Kiess, H.O. (1996). Statistical concepts for the behavioral sciences (2nd ed.). Boston, MA: Allyn and Bacon.
- Knoff, H. M. (1995). Best practices in personality assessment. In A. Thomas & J. Grimes (Eds.), Best practices in school psychology-III (pp. 849–864). Washington, DC: National Association of School Psychologists.
- Landis, J.R., & Koch, G.C. (1977). The measurement of observer agreement for categorical data. Biometrics, 33, 159-174.
- McConaughy, S.H., & Ritter, D.R. (1995). Best practices in multidimensional assessment of emotional or behavioral disorders. In A. Thomas & J. Grimes (Eds.), Best practices in school psychology-III (pp. 865–877). Washington, DC: National Association of School Psychologists.
- McDermott, P.A. (1988). Agreement among diagnosticians or observers: Its importance and determination. Professional School Psychology, 3, 225–240.
- McDermott, P.A. (1994). National profiles in youth psychopathology: Manual of Adjustment Scales for Children and Adolescents. Philadelphia: Edumetric and Clinical Science.
- McDermott, P.A., Marston, N.C., & Stott, D.H. (1993). Adjustment Scales for Children and Adolescents. Philadelphia: Edumetric and Clinical Science.
- McDermott, P.A., Watkins, M.W., Sichel, A.F., Weber, E.M., Keenan, J.T., Holland, A.M., & Leigh, N.M. (1995). The accuracy of new national scales for detecting emotional disturbance in children and adolescents. The Journal of Special Education, 29, 337–354.
- McDermott, P.A., & Weiss, R.V. (1993). Syndromic profile types for the Adjustment Scales for Children and Adolescents. Philadelphia: University of Pennsylvania.
- McDermott, P.A., & Weiss, R.V. (1995). A normative typology of healthy, subclinical, and clinical behavior styles among American children and adolescents. Psychological Assessment, 7, 162–170.
- Merrell, K.W. (1994a). Assessment of behavioral, social, & emotional problems. New York: Longman.
- Merrell, K.W. (1994b). Preschool and Kindergarten Behavior Scales. Brandon, VT: Clinical Psychology Publishing Company.
- Milich, R., Widiger, T.A., & Landau, S. (1987). Differential diagnosis of attention deficit and conduct disorders using conditional probabilities. Journal of Consulting and Clinical Psychology, 55, 762–767.
- Naglieri, J., LeBuffe, P.A., & Pfeiffer, S.I. (1993). Devereux Behavior Rating Scale–School Form. San Antonio, TX: The Psychological Corporation.
- Quay, H.C. (1986). Classification. In H.C. Quay & J.S. Werry (Eds.), Psychopathological disorders of childhood (3rd ed., pp. 1–42). New York: Wiley.
- Reschly, D.J., & Ysseldyke, J.E. (1995). School psychology paradigm shift. In A. Thomas & J. Grimes (Eds.), Best practices in school psychology-III (pp. 17–32). Washington, DC: National Association of School Psychologists.
- Reynolds, C.R., & Kamphaus, R.W. (1992). Behavior Assessment System for Children. Circle Pines, MN: American Guidance Service.

Salvia, J., & Ysseldyke, J.E. (1995). Assessment. Boston: Houghton Mifflin.

Stinnett, T.A., Havey, J.M., & Oehler-Stinnett, J. (1994). Current test usage by practicing school psychologists: A national survey. Journal of Psychoeducational Assessment, 12, 331–350.

Stott, D.H. (1966). The social adjustment of children. London: University of London Press.

Stott, D.H., Marston, N.C., & Neill, S.J. (1975). Taxonomy of behaviour disturbance. Toronto, Ontario: Musson.

Trites, R.L., Blouin, A.G.A., & Laprade, K. (1982). Analysis of the Conners Teacher Rating Scale based on a large normative sample. Journal of Consulting and Clinical Psychology, 50, 615–623.