Interrater Agreement for Syndromic Profile Classifications on the Adjustment Scales for Children and Adolescents

Gary L. Canivez
Eastern Illinois University

Marley W. Watkins
The Pennsylvania State University

This article reports on an investigation of the interrater agreement on the Adjustment Scales for Children and Adolescents (ASCA) syndromic profile classifications. Teaching professionals (N = 29) who shared the same classroom for a minimum of one hour per day provided independent ratings of the same child on the ASCA. Results indicated that statistically significant interrater agreement was achieved across all 22 syndromic profile classification levels. However, good clinical significance was obtained only for the three- and two-level broad classifications while fair clinical significance was obtained for the five-level broad classifications. Thus, there was adequate interrater agreement for ASCA classifications that reflected broad levels of adjustment and maladjustment but inadequate agreement for specific behavioral types of adjustment and maladjustment. Additional research with much larger sample sizes is needed to better determine the agreement for the 22 syndrome profiles.

Current psychological practice shows an increasing preference for objective assessment techniques that facilitates a link between assessment and intervention (Reschly & Ysseldyke, 1995; Piacentini, 1993). For example, standardized behavior rating scales and checklists have become popular among applied psychologists (Hart & Lahey, 1999; Merrell, 1994) and are the most frequently used instruments in assessing emotional and behavioral difficulties in youths among school psychologists (Stinnett, Havey, & Oehler-Stinnett, 1994). Behavior rating scales are efficient and effective (Knoff, 1995) and a "best practice" (McConaughy & Ritter, 1995) in assessing the emotional and behavioral disorders of school children. Teachers are considered to be among the most accurate adult raters of child behavior (Kamphaus & Frick, 1996) and appear to use a normative perspective in rating child behaviors (Piacentini, 1993) due to their observation of many students across time and contexts.

Behavior rating scales, like all tests, must demonstrate acceptable psychometric properties before they can be validly applied in practice. Behavior rating scales differ across a number of psychometric dimensions (Edelbrock, 1983), and one of the most critical psychometric properties of any instrument relying on informants is the degree to which those informants (raters) agree. Commonly referred to as interrater or interobserver agreement, this measures the extent to which conclusions drawn from an instrument vary as a function of the rater, not the student being rated.

According to Martin, Hooper, and Snow (1986), subjectivity of raters is the primary source of error in rating scale data. For example, when assessing a student's emotional and behavioral adjustment, two teachers observing the same student in the identical classroom environment should report similar types and levels of behavior on a rating scale. If they do not, results would not generalize to other raters and could be due to instrument or rater error rather than student behavior. If raters do agree,

scores can be generalized to other raters and, in a theoretical sense, represent the scores of all raters for that student.

The *Adjustment Scales for Children and Adolescents* (ASCA; McDermott, Marston, & Stott, 1993) is a relatively new behavior rating scale designed to assess youth psychopathology in school settings. McDermott (1994) presented a multivariate method of interpretation of the ASCA that is based on results of a cluster analysis of the ASCA standardization sample, which produced 22 distinct profile (behavioral) types (McDermott, 1994; McDermott & Weiss, 1995). This syndromic profile classification involves comparing a youth's core syndrome T score profile to the 22 ASCA profile types to determine the normative profile to which the youth's scores are most similar. Similarity is quantified by the generalized distance score (GDS), which takes into account both level and directional information (Cronbach & Gleser, 1953; Osgood & Suci, 1952). Canivez (1998) and Watkins (1997) automated calculation of the GDS to ensure reliable computation.

Each of the 22 syndrome profile types are described in the ASCA manual according to distinguishing characteristics "based on tests of the standard error of proportional differences corrected for multiple contrasts" (McDermott, 1994, p. 22). Such comparisons may facilitate differential diagnosis and better description of youth behavioral characteristics relative to the general population. Characteristics presented in syndrome profile descriptions include behavioral, cognitive, academic, socioeconomic, racial, gender, developmental, and family.

Given the potential differential diagnostic applications of ASCA syndrome profiles (McDermott, 1994), an assessment of their interrater agreement is needed if they are to be used clinically. Although behavior rating scales have many positive qualities, there are a number of potential threats to their validity such as rater bias (i.e. halo effect, leniency error), rater competency, relevant contact, and rater agreement. Rater agreement is not necessarily a problem if the raters observe in different environments as it is possible that behaviors vary with respect to different environments and the scale may simply measure those differences. However, the present study focused on agreement of teachers and aides observing children in the *same* classroom at the *same* time. Thus, the purpose of the present study was to investigate the degree of interrater agreement on syndromic profile classifications produced by different raters observing the same child in the same environment.

Method

Participants

Participants from the Watkins and Canivez (1997) study of interrater agreement of ASCA syndrome T scores also served as participants in this investigation. Teachers from two school districts in two states were recruited to complete ASCA rating forms on their students. Both districts were located in suburban areas of major cities: one in the Southwest and one in the Midwest. A total of 71 students were identified, whose classroom behaviors were jointly observed for a minimum of one hour each day by two professionals or paraprofessionals who were willing to participate in this study. Raters' job classifications included special education teacher, special education teaching assistant, remedial reading teacher, science teacher, and regular classroom teacher. The most frequent rating pair was a special education teacher and a special education teaching assistant in a self-contained, special education setting (58%). Other observer pairs included classroom teacher-special education teacher (38%) and classroom teacher-remedial reading teacher (4%). In total, there were 29 raters comprising 71 pairs within 24 classrooms in 6 different schools.

Students' racial/ethnic backgrounds, as reported by parents on school enrollment forms, included 80% Caucasian, 10% Hispanic/Latino, 7% Black/African American, and 3% other. The student sample was 66%

male and 34% female, ranging in age from 7 through 17 years, with a median age of 11 years and a mean age of 11.1 years. Students were enrolled in grades 1 through 10 and were involved in a variety of special programs for those students at risk or with disabilities: 44% in learning disability; 29% in emotional disability; 19% in severe language impairment; and 8% in mild mental retardation.

Instrument

The *Adjustment Scales for Children and Adolescents* (ASCA; McDermott et al., 1993) is an objective behavior rating instrument completed by a student's classroom teacher and designed for use with all noninstitutionalized youths ages 5 through 17 (grades K through 12). The ASCA consists of 156 behavioral descriptions within 29 specific situations where teachers may observe student behaviors. Of the 156 items, 97 are scorable for psychopathology and based on factor analyses, singularly assigned to one of six core syndromes (Attention-Deficit/Hyperactive, Solitary Aggressive-Provocative, Solitary Aggressive-Impulsive, Oppositional Defiant, Diffident, and Avoidant) or two supplementary syndromes (Delinquent and Lethargic/Hypoactive). The core syndromes are combined to form two composite indexes: Overactivity (Attention-Deficit Hyperactive, Solitary Aggressive-Provocative, Solitary Aggressive-Impulsive, and Oppositional Defiant syndromes) and Underactivity (Diffident and Avoidant syndromes). Raw scores are converted to normalized T scores based on a nationally representative standardization 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.

Extensive reliability and validity evidence is provided in the ASCA manual (McDermott, 1994). 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. Test-retest reliabilities ($n = 40$) over a 30-school-day interval ranged from .66 to .91 for the six core syndromes and from .75 to .79 for the Overactivity and Underactivity scales. Significant stability (Canivez, 2000; Canivez, Perry, & Weller, 2001) and interrater agreement (Canivez, Watkins, & Schaefer, in press; Watkins & Canivez, 1997) have also been observed in independent samples. 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). Additional construct validity evidence has been presented for the ASCA (Canivez & Bordenkircher, in press; Cavinez & Rains, in press; McDermott, 1995; McDermott & Schaefer, 1996; McDermott & Spencer, 1997) and it has demonstrated good diagnostic accuracy in identifying students with emotional disturbance (McDermott et al., 1995). In general, the psychometric characteristics of the ASCA seem acceptable and meet standards for both group and individual decision-making (Canivez, 2001; Salvia & Ysseldyke, 1995).

Procedure

Independent ratings of the 71 participating students were collected over a four-week period according to ASCA standard administration procedures. The student's primary teacher (special education teacher or regular classroom teacher) was designated as Rater 1 whereas the secondary rater (special education teaching assistant, resource teacher, and remedial reading teacher) was designated as Rater 2.

Syndromic profile classifications were automated (Canivez, 1998) using the generalized distance score (GDS) method recommended in

the ASCA manual (McDermott, 1994). The GDS is a measure of profile similarity based upon deviations of a youth's core syndrome T scores from the average T scores for a specified ASCA profile type. The youth's profile is classified as most similar to the ASCA profile type that results in the smallest GDS.

Interrater agreement was analyzed for several alternative classification methods. First, agreement on classification into the 22 ASCA profile types was calculated. McDermott and Weiss (1995) noted that this typology is hierarchical and suggested that alternative categorizations might be explored. Consequently, agreement on classification into five broader categories was examined next. These five types represent severity of maladjustment, where profile Type 1 is classified as Adjusted, Types 2 through 5 are Adequately Adjusted, Types 6 through 12 are Marginally Adjusted, Types 13 through 18 are At-Risk, and Types 19 through 22 are Maladjusted. Third, given that the Adjusted, Adequately Adjusted, and Marginally Adjusted categories all represent nondisabled functioning, they were collapsed into a single Adjusted designation and compared to the At-Risk and Maladjusted classifications. Finally, the At-Risk and Maladjusted classifications were combined into a Not Adjusted category and compared to the Adjusted category. Table 1 details the collapse of the 22 narrow syndromic profiles into the broader five-, three-, and two-category groupings.

Data Analysis

Unlike the ASCA syndrome and global adjustment scale T scores, syndromic profile classifications are nominal scale variables. When investigating agreement on nominal scale or categorical variables, statistics such as kappa (Cohen, 1960; Fleiss, 1981) should be utilized (McDermott, 1988; Watkins & Pacheco, 2000). Conceptually, coefficient kappa is the proportion of agreement over and above what would be expected by chance. Kappa coefficients between profiles produced

by Rater 1 and Rater 2 were calculated using automated spreadsheet templates (Canivez, 1999) for the 22 syndromic profiles and the five-, three-, and two-category groupings.

Results

Interrater agreement for the 22 syndrome profile classifications and the five, three, and two broad classifications based on the GDS method is summarized in Table 2. As illustrated, of the 71 children rated by two independent raters, 31 received the identical syndromic profile classification by both raters. This resulted in an observed agreement of 44% and a kappa coefficient of .39 (z = 11.32, p < .00001). Interrater agreement for the five broad categories showed that of the 71 children rated by the two independent raters, 47 (66%) were classified into the same broad category for a kappa coefficient of .53 (z = 7.61, p < .00001). Interrater agreement for the three broad categories showed that of the 71 children rated by the two independent raters, 53 (75%) were classified into the same category (kappa = .60, z = 6.87, p < .00001). Finally, interrater agreement for the two broad categories indicated that of the 71 children rated by the two independent raters, 60 (85%) were classified into the same category (kappa = .68, z = 5.74, p < .00001).

Discussion

This was the first study to investigate the interrater agreement of the multivariate ASCA syndromic profile classification method. As noted above, results indicated that the 22 syndromic profile classifications and their resulting five-, three-, and two-level broad classifications all demonstrated statistically significant interrater agreement beyond chance based on kappa coefficients.

Cicchetti (1994) suggested the following clinical significance at various kappa levels: below .40 = poor; .40-.59 = fair; .60-.74 = good; .75-1.00 = excellent. According to these standards, agreement on the 22 syndrome profiles

Table 1 Broad Classifications of the 22 Syndromic Profiles into Five, Three, and Two Classification levels

Syndromic Profile Type Classification	Five-Level Classification	Three-Level Classification	Two-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. Marginal Adjustment w/Withdrawal	Marginal Adjustment	Adjusted	Adjusted
7. Marginal Adjustment w/Motivation Deficit	Marginal Adjustment	Adjusted	Adjusted
8. Marginal Adjustment w/Avoidance	Marginal Adjustment	Adjusted	Adjusted
9. Marginal Adjustment w/Attention Seeking	Marginal Adjustment	Adjusted	Adjusted
10. Marginal Adjustment w/Moodiness	Marginal Adjustment	Adjusted	Adjusted
11. Marginal Adjustment w/Nonparticipation	Marginal Adjustment	Adjusted	Adjusted
12. Marginal 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

Table 2 Stability of ASCA Syndromic Profile-Based Classifications Using the GDS Method

	P_o	P_c	κ
22 Syndrome Profiles	.44	.08	.39[*]
5 Broad Classifications	.66	.28	.53[*]
3 Broad Classifications	.75	.37	.60[*]
2 Broad Classifications	.85	.51	.68[*]

Note. P_o = proportion of observed agreement, P_c = proportion of chance agreement, κ = kappa. Syndromic profile classification agreement analysis tables are available from the first author upon request.
[*]$p < .0001$.

almost attained fair clinical significance. Agreement at the five-profile level was fair, while agreement at the three- and two-profile level achieved good clinical significance. Thus, there was fair to good interrater agreement on ASCA classifications that reflected broad levels of adjustment-maladjustment, but poor to fair agreement on specific behavioral types of adjustment and maladjustment.

This is a somewhat encouraging and important finding to the extent that one would expect a profile generated by a behavior rating scale to be similar for two raters observing the same child in the same classroom. As no previous studies have investigated interrater agreement between nominal scale multivariate interpretive classification methods and behavior rating scales, it is difficult to place the present results in a broader perspective. However, Danforth and DuPaul (1996) found significant interrater agreement for several teacher rating scales used in assessing attention-deficit hyperactivity disorder (*Mdn* kappa = .51). Likewise, the present levels of diagnostic agreement compare favorably to those found in studies of structured interviews (computer versus clinician kappa = .23) for psychiatric diagnoses (Fisher et al., 1997; Hodges & Zeman, 1993). Diagnostic agreement of ASCA profiles also was consistent with kappas (.54 to .59) reported for the DSM-IV field trials for disruptive behavior disorders (Lahey et al., 1994). Thus, the five-, three-, and two-level broad ASCA syndrome profile classifications demonstrated diagnostic agreement adequate for clinical use.

Given the low kappa coefficient for the 22 syndrome profile comparison, diagnostic use of these narrow syndrome profile classifications warrants caution and is not recommended until further reliability and validity studies are conducted. This approach is best considered a descriptive method for better understanding the relationship between the individual student and the general population with respect to behavioral, cognitive, academic, and other characteristics.

In considering the present results it is important to note that the sample size was small and that there may have been too few children per syndrome profile. Some syndrome profiles were not produced (Type 21 and Type 22) by either rater while others were represented by only one student. This may have artificially constricted kappa coefficients in the 22 profile agreement comparison. When collapsing the 22 profiles into fewer broad classifications, this constriction was eliminated, resulting in higher kappa coefficients. Larger sample sizes and a more diverse sample with respect to behavioral functioning may help address this issue.

Caution should also be exercised in interpreting the results of the present study as our

small, nonrandom sample of students was not representative of the population at large. Generalizability may also be circumscribed as this study employed a limited number of raters, classrooms, and geographic locations. Future studies should continue to investigate the inter-rater agreement of the ASCA in a similar manner while incorporating larger and more diverse and representative student and teacher samples. Larger samples would allow for a better test of the 22 profile agreement comparison. Replication within regular education settings is particularly needed as behavior rating scales are frequently used in these settings for screening and initial evaluations to determine psychopathology and disability. However, it is extremely difficult to find regular education classrooms where two teachers are present at the same time.

References

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

Canivez, G. L. (1998). 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. (1999). *Automated Calculation of Nominal Scale Agreement (Kappa) Statistics for Interrater Agreement and Stability Investigations of ASCA Syndromic Profile Classifications and Discriminant Classifications.* Microsoft® Excel™ Spreadsheet Templates for the Apple® Macintosh™ 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. TM030664)

Canivez, G. L. (2001). *Review of the Adjustment Scales for Children and Adolescents.* In J. Impara & B. Plake (Eds.), *The fourteenth mental measurements yearbook* (pp. 22-24). Lincoln: University of Nebraska, Buros Institute of Mental Measurements.

Canivez, G. L., & Bordenkircher, S. E. (in press). Convergent and divergent validity of the *Adjustment Scales for Children and Adolescents* and the *Preschool and Kindergarten Behavior Scales. Journal of Psychoeducational Assessment.*

Canivez, G. L., Perry, A., & Weller, E. (2001). Stability of

the *Adjustment Scales for Children and Adolescents. Psychology in the Schools, 38*, 217-227.

Cavinez, G. L., & Rains, J. (in press). Construct validity of the *Adjustment Scales for Children and Adolescents* and the *Preschool and Kindergarten Behavior Scales:* Convergent and divergent evidence. *Psychology in the Schools.*

Canivez, G. L., Watkins, M. W., & Schaefer, B. A. (in press). Interrater agreement for discriminant classifications for the *Adjustment Scales for Children and Adolescents. Psychology in the Schools.*

Cicchetti, D. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and 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.

Cronbach, L. J., & Gleser, G. C. (1953). Assessing similarity between profiles. *Psychological Bulletin, 50*, 456-473.

Danforth, J. S., & DuPaul, G. J. (1996). Interrater reliability of teacher rating scales for children with attention-deficit hyperactivity disorder. *Journal of Psychopathology and Behavioral Assessment, 18*, 227-237.

Edelbrock, C. (1983). Problems and issues in using rating scales to assess child personality and psychopathology. *School Psychology Review, 12*, 293-299.

Fisher, P., Lucas, C., Shaffer, D., Schwab-Stone, M., Graae, F., Lichtman, J., et al. (1997). *Diagnostic Interview Schedule for Children, Version IV (DISC-IV): Test-retest reliability in a clinical sample.* Poster presented at the 44th annual meeting of the American Academy of Child and Adolescent Psychiatry, Toronto.

Fleiss, J. L. (1981). *Statistical methods for rates and proportions* (2nd ed.). New York: Wiley.

Hart, E. L., & Lahey, B. B. (1999). General child behavior rating scales. In D. Shaffer, C. P. Lucas, & J. E. Richters (Eds.), *Diagnostic assessment in child and adolescent psychopathology* (pp. 65-87). New York: Guilford.

Hodges, K., & Zeman, J. (1993). Interviewing. In T. H. Ollendick & M. Hersen (Eds.), *Handbook of child and adolescent assessment* (pp. 65-81). Boston: Allyn and Bacon.

Kamphaus, R. W., & Frick, P. J. (1996). *Clinical assessment of child and adolescent personality and behavior.* Boston: 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.

Lahey, B. B., Applegate, B., Barkley, R. A., Garfinkel, B., McBurnett, K., Kerdyk, L., et al. (1994). DSM-IV field

trials for oppositional defiant disorder and conduct disorder in children and adolescents. *American Journal of Psychiatry, 151,* 1163-1171.

Martin, R. P., Hooper, S., & Snow, J. (1986). Behavior rating scale approaches to personality assessment in children and adolescents. In H. M. Knoff (Ed.), *The assessment of child and adolescent personality* (pp. 309-351). New York: Guilford.

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. (1994). *National profiles in youth psychopathology: Manual of Adjustment Scales for Children and Adolescents.* Philadelphia: Edumetric and Clinical Science.

McDermott, P. A. (1995). Sex, race, class, and other demographics as explanations for children's ability and adjustment: A national appraisal. *Journal of School Psychology, 33,* 75-91.

McDermott, P. A. (1988). Agreement among diagnosticians or observers: Its importance and determination. *Professional School Psychology, 3,* 225-240.

McDermott, P. A., Marston, N. C., & Stott, D. H. (1993). *Adjustment Scales for Children and Adolescents.* Philadelphia: Edumetric and Clinical Science.

McDermott, P. A., & Schaefer, B. A. (1996). A demographic survey of rare and common problem behaviors among American students. *Journal of Clinical Child Psychology, 25,* 352-362.

McDermott, P. A., & Spencer, M. B. (1997). Racial and social class prevalence of psychopathology among school-age youth in the United States. *Youth & Society, 28,* 387-414.

McDermott, P. A., Watkins, M. W., Sichel, A. F., Weber, E. M., Keenan, J. T., Holland, A. M., et al. (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. (1995). A normative typology of healthy, subclinical, and clinical behavior styles among American children and adolescents. *Psychological Assessment, 7,* 162-170.

Merrell, K. W. (1994). *Assessment of behavioral, social, & emotional problems.* New York: Longman.

Osgood, C. E., & Suci, G. J. (1952). A measure of relation determined by both mean difference and profile information. *Psychological Bulletin, 49,* 251-262.

Piacentini, J. (1993). Checklists and rating scales. In T. H. Ollendick & M. Hersen (Eds.), *Handbook of child and adolescent assessment* (pp. 82-97). Boston: Allyn and Bacon.

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.

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.

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.

Watkins, M. W. (1997). ASCAcalc [Computer Software]. State College, PA: Author.

Watkins, M. W., & Canivez, G. L. (1997). Interrater agreement of the *Adjustment Scales for Children and Adolescents. Diagnostique, 22,* 205-213.

Watkins, M. W., & Pacheco, M. (2000). Interobserver agreement in behavioral research: Importance and calculation. *Journal of Behavioral Education, 10,* 205-212.