

Note on Truncated T Scores in Discrepancy Studies with the Child Behavior Checklist and Youth Self Report

Steven Thurber, Ph.D. and William Sheehan, M.D.

Abstract

There is an extensive literature on discrepancies between adult informant ratings of symptom severity on the Child Behavior Checklist and ratings by children and adolescents on the Youth Self Report. This article presents an elaboration of a caveat against the use of T scores and for employing raw scores in statistical analyses related to such discrepancies. The T score transformation on these instruments is truncated resulting in reduced variability; this, in turn, may affect comparability of T scores with data based on the full range of variation. Children and adolescents (N = 236) responded to the items of the Youth Self Report while their adult caregivers completed the Child Behavior Checklist. Mean difference tests between youth and caregiver ratings as well as logistic regression, were done separately for raw scores and T scores; the results illustrated the importance of following the aforementioned caveat.

Introduction

A well-accepted precept in the assessment of children and adolescents involves gathering psychometric data from several sources in order to ascertain situation-specific as opposed to possible generalization of behavioral and emotional problems across home and school settings (Achenbach, McConaughy, & Howell, 1987). Moreover, it is important to evaluate psychopathology on the basis of two or more data sources in order to evaluate error related to method variance (Campbell & Fiske, 1959).

The approach pioneered by Thomas Achenbach and associates, the Achenbach System of Empirically Based Assessment (ASEBA), includes a checklist of symptoms and factor analytically derived scales, with separate forms for adult caregivers, teachers, and the child or adolescent age 11 through 18. Similar items among the test forms, with content differing only in terms of first versus third person, allow direct comparisons among parents, teachers, and the individual's self report (Achenbach, 1991; Achenbach & Rescorla, 2001).

An important clinical and research issue relates to the dissimilarities that occur in comparisons between adult informants ratings on the Child Behavior Checklist (CBCL) of the ASEBA and the child or adolescent's perspective on the Youth Self Report (YSR). It is possible, in fact fairly common in clinical settings, for a parent to rate her or his offspring as having symptoms in the pathological realm of severity, whereas the youths' ratings are within normal limits (Thurber & Hollingsworth, 1992). Discrepant results between informants are a very robust finding as reported in major reviews of these assessment instruments (Achenbach et al., 1987; De Los Reyes & Kazdin, 2005). There is an extensive literature related to causal analyses

of such disagreements that includes posited emotional dysfunctions of adult informants leading to elevated severity ratings (Friedlander, Weiss, & Traylor, 1986), and “faking good” or minimizing problems on the part of youth (Author & Hollingsworth, 1992).

Several techniques have been devised for dealing with adult-youth disagreement including the use of the CBCL-YSR discrepancy score as a predictor variable in relation to later maladjustment (Ferdinand, van der Ende, & Verhulst, 2006) and analyses with adult and youth reports functioning as method variance in multitrait-multimethod investigations (Thurber & Snow, 1990; Grigorenko, Geiser, Slobodskaya, & Francis, 2010). There is also an extant theoretical model for explaining discrepancies between and among informants (De Los Reyes & Kazdin, 2005).

Adult-youth discrepancies constitute a central arena of research in studies comparing scores on the CBCL and YSR. Researchers in this domain have the choice between raw scores or standardized (T) scores in data analysis. It appears that investigators have been rather unsure as to which type of score is most appropriate. It is fair to say that at least until the turn of the century, raw scores were employed mainly with community samples whereas T scores were most frequently used with clinical samples (see Handwerk, Larzelere, Soper, & Friman, 1999). The latter makes intuitive sense since clinical researchers would show greater interest in a standard score that gives immediate information as to the child or adolescent’s relative standing on these tests of psychopathology.

After the publication of a new manual for the 2001 version of the scales there should be no ambiguity regarding statistical usage of raw or T scores. Achenbach and Rescorla (2001, p. 89) state explicitly that raw scale scores should be used in statistical analysis in order to take account of the full range of variation in these scales. Recall that a T score is an invariant transformation of a Z score (adding a constant of 50 to the mean of zero and 9 to the standard deviation) which in turn is an invariant transformation of a raw score (subtracting the mean constant from each score and dividing by the standard deviation constant).

Researchers using the Achenbach system may not fully understand that the standardized T scores used in the current CBCL and YSR are truncated. This refers to the fact that the T transformation results in the elimination of the bottom part of the score distribution; lower scores at the mean and below are all assigned T scores of 50, not just the mean raw score as in non-truncated T scores. Hence, T scores truncate or reduce the range of variation.

Confusion regarding the use of raw or T scores in statistical analyses still appears extant in recent research published in reputable journals. Some investigators and journal reviewers continue to seem oblivious to issues of raw scores versus T scores. For example, Berg-Nielsen, Vika, and Dahl (2003) (*Child: Care, Health & Development*) used raw scores only because T scores were not available. Leung, Kwang, Tang, Ho, and Hung (2006) employed T scores in measuring change across time, in a study published in the *Journal of Child Psychology and Psychiatry*; Salbach-Andrae, Klinkowski, Lenz, and Lehmkuhl (2009) reported T scores in assessing the degree of agreement between parents and youth, published in the *European Journal of Child and Adolescent Psychiatry*. In the *Child and Family Behavior Therapy* journal, Glaser, Calhoun, Bradshaw, Bates, and Socherman (2011), used T scores in evaluating agreement among raters using the CBCL and YSR. Fung and Lau (2010) found disagreements between parents and children using T scores, published in the *Journal of Clinical Child and Adolescent Psychology*.

The current study is in keeping with the corpus of investigations on CBCL-YSR discrepancies. The aim, however, is to compare the magnitude and order of variables in separate raw and T score analyses in relation to those discrepancies.

Method

Participants. The participants were 236 children and adolescents representing consecutive admissions to a community mental health center who were administered the YSR. There were 100 girls and 136 boys ranging in age from 11 to 18 years old ($M = 14$ years 2 months, $SD = 2$ years 1 month). There were 194 Caucasians, 23 from Somalia, 3 Native Americans, 3 African Americans, and 16 individuals with unknown ethnicity. The adult caregivers of these individuals (87 women, 149 men) completed the CBCL. Upon arrival at the center, the child or adolescent and accompanying adult met with a mental health professional and completed the YSR and CBCL independently. A consent form was signed indicating agreement for using the obtained results for diagnostic and treatment considerations and for research in which anonymity would be maintained.

Measures. The aforementioned CBCL and YSR have DSM-oriented scales rationally derived from DSM criteria, and factor-analytically developed syndrome or problem scales. The latter include Anxiety/Depression, Withdrawn/Depressed, Somatic Problems, Social Problems, Thought Problems, Attention Problems, Rule Breaking, and Aggression. Since the DSM scales have not been found to yield incremental utility beyond the empirical problem scales, only the latter were included in data analysis (Ebesutani, Bernstein, Nakamura, Chorpita, Higa-McMillan, & Weisz (2010).

Procedure. The data from the CBCL and YSR were subjected to mean comparisons and logistic regression using IBM SPSS 20 (2011). Separate analyses were conducted for raw scores and T scores.

Results

Inspection of the data in Table 1 reveals that virtually all T score means for adult informants are elevated (i.e., more severe in magnitude) when compared to the child and adolescent self ratings. Only the “Somatic Problems” mean fails to attain traditional standards of statistical reliability. In contrast, when the full levels of variability are obtained (see raw scores in Table 2), both “Somatic Problems” ($p < .0005$) and “Thought Problems” ($p < .003$) are significantly elevated for the Youth ratings. With reference to poorly controlled symptoms (Externalizing), “Attention Problems,” “Rule Breaking,” and “Aggression” are significantly elevated in the CBCL ratings regardless of raw or T score scaling (all p values $< .05$). Conclusions for T scores only would disguise elevation for the more private somatic and thought difficulties.

Table 1

Means and Standard Deviations for T Scores

Scale	Adult Informant <i>M (SD)</i>	Child/Adolescent <i>M (SD)</i>	<i>p</i>
Anxiety/Depression	64.322 (10.770)	61.470 (10.689)	.004
Withdrawn/Depressed	66.663 (11.217)	61.339 (9.646)	.0005
Somatic Problems	60.975 (9.916)	59.661 (9.195)	.136
Social Problems	63.797 (9.503)	61.127 (9.545)	.002
Thought Problems	63.691 (8.861)	60.360 (10.270)	.0005
Attention Problems	64.703 (9.602)	62.322 (11.166)	.031
Rule Breaking	63.963 (8.815)	59.233 (8.044)	.0005
Aggression	66.627 (11.624)	60.631 (9.075)	.0005

Table 2

Means and Standard Deviations for Raw Scores

Scale	Adult Informant <i>M (SD)</i>	Child/Adolescent <i>M (SD)</i>	<i>p</i>
Anxiety/Depression	7.771 (5.327)	7.932 (5.854)	.755
Withdrawn/Depressed	6.610 (3.741)	5.525 (3.320)	.795
Somatic Problems	3.661 (3.835)	4.932 (3.852)	.0005
Social Problems	5.911 (4.197)	5.936 (4.141)	.947
Thought Problems	5.166 (3.811)	6.398 (5.096)	.003
Attention Problems	8.953 (4.312)	8.072 (3.948)	.021
Rule Breaking	7.799 (5.357)	6.648 (4.626)	.023
Aggression	13.137 (8.327)	10.636 (5.755)	.0005

Table 3 includes information on the Wald statistical test for the logistic regression coefficients, with reference to the relative strengths of the problem scales regressed on the adult-youth dichotomy. Also note the inclusion of separate rank orders for raw and T scores. The Kendall tau rank order correlation between these values was -0.286 (ns). Using truncated T scores would yield radically different interpretations in comparison to interpretations based on the full range of variability.

Table 3

Wald Measure Comparisons for Raw Scores and T Scores

Scale	Raw Scores	Rank	T Scores	Rank
Anxiety/Depression	3.635	5	.006	8
Withdrawn/Depressed	.801	7	4.337*	2
Somatic Problems	10.304**	3	.689	5
Social Problems	1.953	6	.024	7
Thought Problems	18.013***	1	.550	6
Attention Problems	5.985*	4	1.771	4
Rule Breaking	.014	8	10.668**	1
Aggression	14.156***	2	2.076	3

Note. * $p < .05$. ** $p < .001$. *** $p < .0001$.

Discussion

The impetus for the current investigation was the premise that raw scores and truncated T scores are incompatible. Hence, published studies that vary regarding the use of raw scores and T scores cannot be integrated. Furthermore, conclusions based on reduce variability (truncated T scores) may be untenable. Even though Achenbach and Rescola (2001) clearly advocate employing raw scores in data analysis, published studies continue to use T scores in their analyses. We posited that a presentation on the meaning of truncation together with data illustrating the statistical effects of eliminating the lower portion of the score distribution would raise awareness and increase the use of raw scores for data analysis in future studies. It seems clear from the current study that researchers should employ the full range of variation in data analysis. Without acceding to this caveat, truncated data in effect are “disturbed.” For our current data, truncated T scores had the effect of disguising the level of symptom severity of youths. In comparison to adult ratings, for example, T score means for youths were spuriously low on certain internalizing, private symptoms involving somatic and thought difficulties. These same T

scales were nonsignificant contributing variables in regression analysis but were highly significant as raw scores in logistic regression. When the distribution of scores on any psychological measure is artificially attenuated, statistical reliabilities will likely be different when compared to statistical tests using a full range of data. This is simply a matter of the nature of variability in relation to statistical analyses, learned in a first course in statistics.

In the current study, truncated comparisons suggested essentially greater severity of general symptomatology from the adult perspective. This was likely a distorted picture. The distortion is evident when a full range of data are considered. With non-truncated raw scores, the differences between adult and child/adolescent CBCL-YSR ratings are markedly changed for Internalizing symptoms. Differences reflecting more elevated ratings by adults are either eliminated or, in two instances, represented more severe self ratings by youths.

There is always a question regarding generalizations of findings. In this study, data were gathered from participants who would likely evince a fair range of scores on the CBCL and YSR; from low to moderate in an outpatient service and from moderate to severe among individuals in partial hospitalization. The results on Internalizing scales may be unique to these types of clientele. However, these findings with respect to Internalizing symptoms may have ecological validity since they may not be as sensitive to social desirability as youth ratings on Externalizing scales (Thurber & Hollingsworth, 1992).

It is unknown whether the outcomes of this research will extrapolate to persons in different geographic and cultural localities or to individuals in areas with different degrees of urbanization. Nevertheless, the caveat of Achenbach and Rescorla (2001) is congruent with principles of variability and statistical inference and is confirmed by the current investigation.

About the Authors

Steven Thurber, Ph.D. is with the, Department of Psychology, Woodland Centers
Correspondence concerning this article should be addressed to Steven Thurber, Department of Psychology, Woodland Centers, 1125 SE Sixth Street, Willmar, Minnesota 56201.
E-mail: steven_thurber@yahoo.com

William Sheehan, M.D. is a psychiatrist in the neurodevelopmental disorders program at the Child and Adolescent Behavioral Health Services Hospital in Willmar, Minnesota
E-mail: william.p.sheehan@state.mn.us

References

- Achenbach, T. M. (1991). *Integrative guide for the 1991 CBCL/4-18, YSR and TRF Profiles*. Burlington: Department of Psychiatry, University of Vermont.
- Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, *10*, 213-232.
- Achenbach, T. M., & Rescorla, L. (2001). *Manual for the ASEBA School?-Age Forms & Profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, and Families.
-

- Berg-Nielsen, T. S., Vika, A., & Dahl, A. A. (2003). When adolescents disagree with their mothers: CBCL-YSR discrepancies related to maternal depression and adolescent self-esteem. *Child: Care, Health and Development, 29*, 207-213.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.
- Ebesutani, C., Bernstein, A., Nakamura, B. J., Chorpita, B. F., Higa-McMillan, C. K., & Weisz, J. R. (2010). Concurrent validity of the Child Behavior Checklist DSM-Oriented scales: Correspondence with DSM diagnoses and comparison to syndrome scales. *Journal of Psychopathology and Behavioral Assessment, 32*, 373-384.
- De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin, 131*, 483-509.
- Ferdinand, R. F., van der Ende, J., & Verhulst, F. C. (2006). Prognostic value of parent-adolescent disagreement in referred sample. *European Journal of Child and Adolescent Psychiatry, 15*, 156-162.
- Friedlander, S., Weiss, D. S., & Traylor, J. (1986). Assessing the influence of maternal depression on the validity of the Child Behavior Checklist. *Journal of Abnormal Psychology, 14*, 123-183.
- Fung, J. J., & Lau, A. S. (2010). Factors associated with parent-child (dis)agreement on child behavior and parenting problems in Chinese immigrant families. *Journal of Clinical Child & Adolescent Psychology, 39*, 314-327.
- Glaser, B. A., Calhoun, G. B., Bradshaw, C. P., Bates, J. M., & Socherman, R. E. (2011). Multi-observer assessment of problem behavior in adjudicated youths: Patterns of discrepancies. *Child & Family Behavior Therapy, 23*, 33-45.
- Grigorenko, E. L., Geiser, C., Slobodskaya, H. R., & Francis, D. J. (2010). Cross-informant symptoms from CBCL, TRF, and YSR: Trait and method variance in a normative sample of Russian youths. *Psychological Assessment, 22*, 893-911.
- Handwerk, M. L., Larzelere, R. E., Soper, S. H., & Friman, P. C. (1999). Parent and child discrepancies in reporting severity of problem behaviors in three out-of-home settings. *Psychological Assessment, 11*, 14-23.
- IBM SPSS Statistics 20. (2011). Armonk, New York: IBM Corporation.
- Leung, P. W. L., Kwong, S. L., Tang, C. P., Ho, T. P., Hung, S. F., Lee, C. C., Hong, S. L., Chiu, C. M., & Liu, W. S. (2006). Test-retest reliability and criterion validity of the Chinese version of CBCL, TRF, and YSR. *Journal of Child Psychology and Psychiatry, 47*, 970-973.
-

- Salbach-Andrae, H., Klinkowski, N., Lenz, K., & Lehmkuhl, U. (2009). Agreement between youth-reported and parent-reported psychopathology in a referred sample. *European Child and Adolescent Psychiatry, 18*, 136-143.
- Thurber, S., & Snow, M. (1990). Assessment of adolescent psychopathology: Comparison of mother and daughter perspectives. *Journal of Clinical Child Psychology, 19*, 249-253.
- Thurber, S., & Hollingsworth, D. K. (1992). Validity of the Achenbach and Edelbrock Youth Self-Report with hospitalized adolescents. *Journal of Clinical Child Psychology, 21*, 249-254.
-