Published in final edited form as:

J Clin Child Adolesc Psychol. 2009 March ; 38(2): 263–272. doi:10.1080/15374410802698446.

Conners' Teacher Rating Scale for Preschool Children: A Revised, Brief, Age-Specific Measure

David J. Purpura and Christopher J. Lonigan

Department of Psychology, Florida State University.

Abstract

The *Conners' Teacher Rating Scale--Revised* is one of the most commonly used measures of child behavior problems; however, the scale length and the appropriateness of some of the items on the scale may reduce the usefulness of the CTRS for use with preschoolers. In this study, a Graded Response Model analysis based on Item Response Theory was applied to the CTRS data from 669 preschool children. Children in this data sample ranged in age from 25 to 74 months, 44.4% were identified as African American, 45.4% Caucasian, 10.2% other. The sample was 53.7% boys and 46.3% girls. A revised version of the scale was developed specifically for use with preschool children. Five items each for the Inattention, Hyperactivity/Impulsivity, and Opposition scale were included in the revised scale. The revised scales significantly reduce the time needed for teachers to complete the measures while retaining the scales' ability to discriminate children with different levels of behavioral problems.

Keywords

CTRS; ADHD; IRT; assessment; behavior

Item Response Theory Analysis of the Conners' Teacher Rating Scale for Preschool Children

Early behavioral intervention may be crucial when it comes to reducing the negative impact of behavioral problems. To treat behavioral problems successfully, it is necessary to assess such problems effectively. Accurately measuring behavioral problems allows researchers to target treatment development to the specific behavioral problem, and allows clinicians to have a clear picture of any relevant behavioral problems. Additionally, effective assessment of behavioral problems enables researchers to include strong measures of behavioral problems in their research. Current methods of measuring behavioral problems are often time-consuming because of the length of the measures. It is important that these assessment tools not only accurately measure the spectrum of behavioral problems, but that they are also easy to use by both researchers and individuals such as teachers and classroom observers, who must complete the forms. Further, these measures should be age appropriate as behavioral problems may manifest differently across time (Willoughby, 2003).

Behavioral Problems

Attention Deficit/Hyperactivity Disorder (ADHD) is one behavioral problem that has received significant attention in both classroom and research settings because of its potentially adverse effects on academic outcomes and classroom functioning (American

Psychiatric Association, 2000). The two primary domains of behavior that define ADHD are inattention and hyperactivity/impulsivity. Approximately 1.5% of girls and 3% of boys in preschool display clinically significant ADHD symptoms (Cuffe, Moore, and McKeown, 2005). Although the diagnostic stability of an ADHD in preschool children is high after three years (Lahey, Pelham, Loney, Kipp, Ehrhardt, Lee, et al., 2004), the specific subtype diagnosis is unstable over the same time period, indicating that it may be best to examine these domains on a continuum, rather than in diagnostic categories (Lahey, Pelham, Loney, Steve, & Wilcutt, 2005).

Other behaviors associated with ADHD, such as oppositional behaviors, often pose a challenge in the classroom because of their adverse effects on classroom functioning. Disruptive and defiant attitudes and actions can hinder class activities and make learning difficult for both the child with disruptive behaviors and other children in the classroom. The time teachers spend focusing on correcting disruptive and oppositional behaviors takes away from time that could be spent achieving academic goals.

Conners' Teacher Rating Scale

The *Conners' Teacher Rating Scale* (CTRS; Conners, 1969, 1997a, 1997b) is a commonly used measure of behavioral problems associated with ADHD, which originally was developed as a measure of behavioral change for pharmacological studies (Conners, 1969). The primary version of the CTRS is a 59-item form (CTRS-R:L; Conners, 1997a) that is comprised of six scales (Oppositional, Cognitive Problems/Inattention, Hyperactivity, Anxious-Shy, Perfectionism, and Social Problems) as well as an 'ADHD Index' score, a 'DSM-IV: Inattention' score, and a 'DSM-IV: Hyperactivity' score. Several abbreviated forms exist including a 28-item short form (Goyette, Conners, & Ulrich, 1978) and a revised short-form (CTRS-R:S; Conners, 1997b), which is also 28 items (sixteen overlap with the original short-form). The CTRS-R:S has three scales (Oppositional, Cognitive Problems/Inattention, and Hyperactivity) and an 'ADHD Index' scale. Two other measures are the IOWA Conners (Pelham, Milich, Murphy, & Murphy, 1989), which is 10 items long and has two scales (Inattention/Overactivity and Aggression) and the Abbreviated Symptom Questionnaire (Sprague & Sleator, 1973), which also has two scales (Restless/Impulsive and Emotional Lability; Parker, Sitarenios, and Conners, 1996).

Measurement Concerns

As the use of the CTRS has become more common in classroom and research settings, structural and logistical problems in the effective use of the measure have become more salient. The three major areas of concern are the factor structure of the measure, the item composition within each factor, and the length of the CTRS. In utilizing a measure of behavioral problems for research, it is important that the factor structure is clearly defined and appropriate for measuring the targeted constructs. Inattention, Hyperactivity/Impulsivity, and Oppositional behaviors are the three factors that are the most logical and necessary factors to measure because they directly map onto the ADHD and oppositional spectrum of behaviors in the Diagnostic and Statistical Manual of Mental Disorders 4th Edition--Text Revision (DSM-IV-TR; APA, 2000). The CTRS-R:L and the CTRS-R:S both contain scales that are intended to measure Inattention, Hyperactivity, and Oppositional behavior. However, the other two brief measures (ASQ and IOWA-Conners) contain scales that are not clear measures of DSM-IV-TR disorders and thus, limit their use in research.

The item composition of each of these three primary scales, specifically regarding the Inattention scale, is a key concern when using these measures for assessing behavioral problems in preschool children. Several of the items on the inattention scales of the CTRS-R:L and CTRS-R:S, are measures of academic competence that *may* be related to inattention

in children at older ages. Specifically, three of these items (“poor in spelling,” “not reading up to par,” and “poor in arithmetic”) are not appropriate measures of any behavioral problems in a preschool population (Gerhardstein, et al., 2003) and lack face validity as measures of inattention. Thus, the CTRS-R:L and the CTRS-R:S inattention scales lack validity for use with preschool children.

The third major concern with the current versions of the CTRS is the length. With the significant increase in early education research over the last 15-20 years (Administration on Children, Youth and Families, 1997; NICHD, 2000) and the accumulation of research indicating that behavioral problems are negatively associated with academic functioning (Hinshaw, 1992; Spira & Fischel, 2005), it is necessary to measure behavioral problems in this type of research. However, the addition of behavioral measures to assessment batteries can be difficult due to teacher time-shortages and cost. The long version of the CTRS typically takes 15 to 20 minutes to complete and the short form typically takes 5 to 10 minutes to complete. In a preschool classroom of 10 children, it would take a preschool teacher approximately between 2.5 hours and 3.5 hours to complete the long-form of the CTRS and approximately 1 hour and 1.5 hours to complete the short-form. Due to teacher shortages and extensive teaching requirements, the workday for teachers is limited in non-scheduled time (Darling-Hammond, 2000). Having a small amount non-scheduled time leaves little room for teachers to spend time formally evaluating children’s behavior by completing lengthy forms for screening or research purposes. When teachers do have the time to complete these measures for research purposes, they must be compensated for their time. The already high costs of conducting large-scale early education research often limit the additional funding that can be spent on behavioral measures. These concerns often result in many researchers opting not to include the behavioral assessment in their study. Not collecting this type of data may leave significant gaps in the empirical knowledge of children’s behavior and its relation to their academic development.

To conduct high quality research, it is necessary not only to collect complete data sets, but also to ensure that measurement of the constructs are valid in all populations being assessed. These item content and length concerns can be resolved through the application of Item Response Theory (IRT) to the CTRS, both ensuring that items provide the maximum amount of information concerning the three behavioral domains and reducing the length of the scale without a the loss of construct relevant information.

Item Response Theory

An empirical analysis of the structure and content of the CTRS is needed to arrive at a revised version of the scale that maximizes discriminating power in each behavioral domain. IRT, introduced by Lord and Novick (1968a), is a tool that is well-suited to such research. IRT is a model-based method of latent trait measurement that relates the amount of individual’s latent ability or attribute to the probability of endorsing an item (Hambleton, Swaminathan, & Rogers, 1981). IRT allows researchers to select items based on item-level characteristics (difficulty of the item and the items ability to discriminate between individuals with latent trait scores above and below the item difficulty). Specifically, multidimensional item response theory, which is used to model the relationships between two or more unobservable variables within one measure, is necessary to examine the structure of the CTRS.

In Classical Test Theory (CTT), the true score of the entire test is used to define ability; however, with IRT, the individual responses and item functioning are used to define ability. The term “ability” is typically used to refer to the “latent traits” being studied, largely because of the extensive use and development of IRT within the educational field. Development of assessment measures using CTT has limited effectiveness because it is

population dependent, and the standard error of measurement is averaged across the spectrum of ability levels (Hambleton, Swaminathan, & Rogers, 1981). IRT not only eliminates these limitations, but also informs the test constructor as to the most optimal way to maximize the discriminating power of a measure across the full range of behavior.

Parameters within an IRT analysis allow selection of items with high discriminating power across a range of the latent trait that is measured. When guessing is not factored into response patterns, IRT models have parameters for both individual person ability and item functioning. The parameters that describe item functioning (assuming no guessing) are referred to as the “*a*” parameter and the “*b*” parameter. The “*b*” parameter is also called the “difficulty parameter” because it measures the difficulty of an item along the ability continuum. In a four option Likert scale, such as the CTRS, there are three “*b*” parameters; *b*₁, *b*₂, and *b*₃. These are the points along the ability spectrum at which a specific response option would be endorsed fifty percent of the time for an individual with a given ability. The “*a*” parameter is also called the “discrimination parameter” because it is a measure of how well an item differentiates between individuals above and below that point on the ability continuum. Elevated “*a*” values suggest items have strong discriminating abilities at a given point along the behavioral spectrum.

Through the use of these parameters, Item Information Functions (IIF), which show the amount of information obtained from an individual item at all points across the spectrum of behavior, can be developed. IIFs show the ability of an item to provide information across the behavioral spectrum. The sum of the IIFs from a test is the Test Information Function (TIF). Two different types of tests can be developed using IIFs and TIFs. The first type of test is one in which categorizing individuals is desired, such as for screening measures. Items with a narrow range of difficulty scores, surrounding the point at which group discrimination is desired, are selected for this type of measure. Items with a range of difficulty parameters can be selected to develop a broad abilities test, one with which it is possible to measure comprehensively a specific latent trait over a large range of ability. This type of test is predominantly used when the ability of interest is dimensional in nature such as measuring children’s behavior.

Current Study

Current versions of the CTRS are inappropriate for use with preschool-age children in classrooms because of the factor structure of the measure, the item composition of one or more scales, or the length of the measure. Changes in the use of the CTRS have dictated that changes be made in its structure to facilitate ease of use as well as to increase its effectiveness. Therefore, the goal of this study was to construct a measure of behavioral problems (Inattention, Hyperactivity/Impulsivity, and Oppositional behaviors) most closely associated with DSM-IV-TR behavioral problems that was brief, psychometrically sound, and appropriate for use with preschool children. This study was designed to achieve this goal through the following three-step process: First, to estimate the information each item provided to its respective scale. Second, to remove any items that did not provide adequate information about their respective behavioral scales. Third, to select five items per behavioral domain that maximized the discriminating power of each of the scales across the full range of behavior.

Method

Participants

Data were collected in public and private preschools serving children from low- to upper middle-socioeconomic statuses as part of two larger studies. Children ranged in age from 25

months to 74 months, with a mean age of 51.35 months ($SD = 8.52$ months). Of the 669 participants, 44.4% were identified as African American, 45.4% Caucasian, 10.2% other. The sample was 53.7% boys and 46.3% girls. Parental consent was obtained for each child prior to the start of the assessment.

Measures

Conners' Teacher Rating Scale—Classroom teachers reported their observations of children's behavior, using a 44-item hybrid version of the *Conners' Teacher Rating Scale* (CTRS). The hybrid version was constructed by combining the 28 items from the CTRS-R:S with the non-overlapping items from the original CTRS short-form (for a more detailed description of the formation of this version, see Gerhardstein, et al., 2003). For preschool children, the hybrid CTRS yields a three-factor model of behavior: Inattention, Hyperactivity/Impulsivity, and Opposition. Teachers rated how often a child exhibited the behavior indexed by each item on a scale of 0 (not at all) to 3 (frequently). Three items on this measure (Items 29, 34, and 41) are related to academic performance ("poor in spelling", "not reading up to par", and "poor in arithmetic") and do not load on any of the three factors (Gerhardstein et al., 2003). These three items were not included in any data analysis.

ADHD Rating Scale-IV: School Checklist—The ADHD Rating Scale-IV: School Checklist was completed by an intervention instructor for a sample of 268 children from one of the larger projects. These instructors were individuals who were either working towards or had completed an undergraduate degree in psychology, communication disorders, special education, or a related field. These instructors spent 20 minutes per day, five days per week, for a period of three months, with each child in small group settings. This measure assesses each of the 18 DSM-IV-TR criteria for ADHD (9 inattention items, 6 hyperactivity items, and 3 impulsivity items). Intervention instructors rated how often a child exhibited the behavior indexed by each item on a scale of 0 (never or rarely) to 3 (very often).

Data Analysis Procedure

A multidimensional IRT model was constructed to assess the item functioning within the context of the three correlated scales. Other dimensions of model fit were not assessed because much of the data used in this study also were used in Gerhardstein et al. (2003) in which the three-factor model was found to be the best-fitting model. Item parameters for the 41 CTRS items were calculated using Mplus version 4.1 (Muthen, & Muthen, 2005).

Five items from each scale were targeted for inclusion in the revised CTRS. Item selection was conducted based on a four-step procedure set forth by Lord (1977). These steps were, (1) select the shape of the desired TIF, (2) select items that contribute unique information to the TIF, (3) calculate the TIF, and (4) compare the actual TIF to the target TIF and if they are not comparable, repeat steps 2 and 3. To accomplish the first step, a broad abilities test was selected. In classroom settings, children's behavioral problem levels may span the entire spectrum of behavioral problem functioning. In the second step, items for each scale were organized into five groups based on their b parameters, and the item that was best able to discriminate between individuals of different ability levels was selected. The third step was accomplished by utilizing Samejima's graded response model (Samejima, 1969), an IRT model specifically designed for Likert type scales, using MODFIT (Stark, 2001). In the fourth step, if the actual TIF did not compare adequately to the target TIF then one or more items were replaced with other items that had not been selected during the second step. Item replacement and TIF comparison was repeated until the actual and target TIF were maximally comparable.

After test construction was complete, it was necessary to compare the revised measure to other, more established measures of behavioral problems. The three behavioral scales were compared to their respective scales on the hybrid CTRS. Because the hybrid version of the CTRS included all the items from the CTRS-R:S, scores for that measure were also compared to the revised CTRS. Additionally, a sample of the children in this study had also been rated on the inattention and hyperactivity/impulsivity factors of the ADHD Rating Scale. These scores were also compared to the revised CTRS. Missing data for these comparisons were accounted for through multiple imputation using NORM Version 2.03 (Schafer, 2000).

Results

Item parameters for each of the 41 items are shown in Table 1. Four Inattention items (Items 9, 16, 17, 21), one Hyperactivity/Impulsivity item (Item 20), and one Opposition item (Item 6) marginally fit their respective scales (i.e., had a parameters below .750) and were not used in item selection. TIFs for the hybrid CTRS are shown in panel 1 of Figure 1. The hyperactivity/impulsivity scale provided almost twice as much information at its peak than the other two scales, because it is comprised of more items. All three scales had Standard Error (SE) values below .5, which suggests that the hybrid CTRS provided adequate information across the ability continuum (Hambleton et al., 1981).

As can be seen in Table 1, several items within each scale had comparable a and b parameters. These results suggested that more items could be removed from the scales without reducing the ability of the scale to discriminate across the behavioral spectrum. Using the previously mentioned four steps of test construction, a short version of the CTRS for preschool children was constructed. Five items were selected for each scale based on the unique information they provided to the scale as well as their ability to discriminate individuals at a variety of levels of behavior. The TIFs formed by each of the three 5-item scales equated to the target TIF (see panel 2 of Table 1). The standard error of all three revised scales was less than .5, which is in the acceptably low range, across the spectrum of behavior (Hambleton, et al., 1981).

The Inattention scale was comprised of 11 items from the hybrid CTRS. Of the 7 well-fitting items, five items that provided unique discrimination abilities were selected for the final behavioral scale. Three items (Items 23, 26, and 39) each had comparable b parameters however; item 26 was the best fit to the revised scale, and it provided more unique information at the upper end of the behavioral spectrum than did the other two items. The other four items selected provided adequate information in the lower and middle ranges of the behavioral spectrum. The final inattention scale was comprised of Items 18, 24, 26, 35, and 44.

Nineteen items from the hybrid CTRS formed the Hyperactivity/Impulsivity scale. The 18 well-fitting items were separated into five groups based on their b parameters. The first group, lowest b parameters, consisted of only Item 1. This item was well suited for inclusion in the final item scale. The second group (Items 3, 7, 8, 19, 30, 33, 38, and 42) was equivalent in terms of a parameters. However, Item 30 provided the most unique information for the overall needs of the five item scale. In group three (Items 2, 13, 14, and 37), Item 13 provided the most unique information to the scale. In the fourth group (Items 15, 28, and 43), and Item 43 was the best fitting item for the scale. The fifth group was composed of Items 32 and 40. Both items had comparable b parameters, but Item 40 had a higher a parameter and was the better overall item for the scale. The revised Hyperactivity/Impulsivity scale was comprised of Items 1, 13, 30, 40, and 43.

Eleven items from the hybrid CTRS formed the Opposition scale. The 10 well-fitting items were separated into five groups based on their b parameters. The first group, the one with the lowest b parameters, consisted of only Item 10. This item was well suited for inclusion in the final item scale. In the second group (Items 4 and 11), Item 11 provided the most unique information for the overall needs of the five item scale. In group three (Items 12 and 25), item 25 provided the most unique information to the scale and also had greater potential for discrimination than Item 12. In the fourth group (Items 5, 22, 27, and 36), Item 22 was the best fitting item for the scale. Item 31 was the only item in the fifth group, and it provided sufficient unique information to be included in the scale. The revised Oppositional scale was comprised of Items 10, 11, 22, 25, and 31.

The revised 15-item CTRS is shown in Appendix A. Items 1-5 represent Hyperactivity/Impulsivity; Items 6-10 represent Inattention; and Items 11-15 represent Opposition. The revised CTRS scales were compared to the CTRS-R:S scales, the original hybrid CTRS scale, and the ADHD Rating Scale-IV. Factor correlations between the scales are presented in Table 2. The 5-item scale for Hyperactivity/Impulsivity was significantly correlated with the other two Hyperactivity/Impulsivity scales (CTRS-R:S, $r = .94, p < .001$; Hybrid CTRS, $r = .94, p < .001$). The 5-item scale for Inattention was significantly correlated to the other two Inattention scales (CTRS-R:S, $r = .32, p < .001$; Hybrid CTRS, $r = .92, p < .001$). The 5-item scale for Oppositional behavior was significantly correlated to the other two Oppositional scales (CTRS-R:S, $r = .91, p < .001$; Hybrid CTRS, $r = .96, p < .001$).

Due to the significant difference between correlations on the Inattention measures, it was necessary to compare them both to an independent measure of inattention (*ADHD Rating Scale-IV: School Checklist*). The 15-item CTRS inattention scores were significantly correlated to the ADHD inattention criteria ($r = .18, p < .001$) but, the CTRS-R:S inattention scores were not significantly correlated to the ADHD inattention scale ($r = .06, p = .310$ for the CTRS-R:S). The Hyperactivity/Impulsivity scales were also compared to the *ADHD Rating Scale-IV: School Checklist* Hyperactivity/Impulsivity scale. Both the 15-item CTRS scale ($r = .22, p < .001$) and the CTRS-R:S scale ($r = .24, p < .001$) were significantly correlated with the ADHD Rating Scale. Tests for differences in correlated correlations were examined, and they indicated no significant differences between the scales, indicating that both the revised scale and the CTRS-R:S scale comparably correlate with the DSM-IV-TR hyperactivity/impulsivity criteria.

Lastly, because the sample population was not racially representative of a national sample, a differential item functioning (DIF) analysis was necessary to ensure that the items did not have variable difficulty parameters as a function of ethnicity. The items on the hybrid CTRS were used for this analysis. A DIF model using the mean DIF as a reference was utilized in this analysis. No significant DIF was found for any items on any of the scales.

Discussion

The primary goal of this study was to develop a revised CTRS (CTRS-15) for preschool children that would enhance research and screening capabilities. Utilizing IRT, five items for each of the three behavioral subscales were selected based on each item's unique contribution to its respective TIF. The revised hyperactivity/impulsivity and oppositional behavior scales performed comparable to other well established scales. The revised inattention scale was more highly correlated to the DSM-IV-TR inattention criteria for than was the CTRS-R:S inattention scale. The revisions made to the CTRS have resulted in a brief, psychometrically sound, measure of behavioral problems that is appropriate to use with preschool children.

Past attempts to develop a brief measure of behavioral problems (CTRS-R:S; Conners, 1997b; Goyette, Conners, Ulrich, 1978; Pelham, Milich, Murphy, & Murphy, 1989; Sprague & Sleator, 1973), have lacked either appropriate scale structures or item content for young children. In this study, the three behavioral scales on the CTRS-15 mapped directly on to behavioral domains found in the DSM-IV-TR and the specific items selected for inclusion reflected those scales appropriately for preschool children. The direct relation between the CTRS-15 and the DSM-IV-TR domains enables researchers to integrate their research more efficiently to the existing literature than could be done with other measures.

Implications for Research, Policy, and Practice

The efficient nature of the CTRS-15 makes inclusion of this measure in large-scale classroom studies more practical. The shorter amount of time necessary to complete this measure, compared to other comparable measures, may reduce teachers' apprehension towards participating in studies that include behavioral measures. Further, the brevity of this measure, coupled with the removal of minimally informative questions from the scale, may help to reduce the amount of missing or incomplete data. These scales enable researchers to examine the effects of each behavioral domain on other variables, such as academic performance, as well as the effects of other variables on each behavioral domain in an efficient manner. Additionally, due to this measure's direct relation to DSM-IV-TR behaviors, teachers could use the CTRS-15 as a screening tool to refer children with potential behavior problems for further evaluation.

Limitations

Prior to formal implementation of the CTRS-15, a few concerns may need to be addressed. Although the purpose of this study was to develop a psychometrically sound behavioral assessment measure for preschool children, it may be wise to examine whether items on the CTRS vary in both difficulty and discrimination properties as a function of age. It may become evident that one item that is a good measure of inattention in preschool children may not be as informative at older ages. The same is also true in the reverse direction, where an item which was a very poor measure of inattention in preschool, such as one related to academic performance, may be much more informative as high academic performance becomes more expected of children. Whereas inattention may negatively impact academic performance, such items are indirect indicators of the construct of interest. Moreover, there are many possible reasons for poor academic performance, only one of which may be problems of attention. It is necessary to examine the developmental continuum of behavioral problems, as it is evident behavioral problems are malleable (Anastopoulos, Shelton, DuPaul, & Guevremont, 1993) and change with age (Lahey et al. 2004). Research that specifically investigates the CTRS item properties for older children would be a valuable asset to the area of behavioral measurement after the preschool period.

A second area that must be addressed is the potential for items to function differently without the presence of other items from the original scale. That is, the CTRS-15 should be validated on another sample of preschool children to ensure that the items function in the same manner as demonstrated in this study. Based on IRT theories of local item independence, it is understood that the items on rating scales are generally independent from influence by other items (Lord & Novick, 1968b); however, it would be prudent to validate the supposed independence of the selected items. Future research can minimize the impact of both areas of limitation through conducting further use of IRT models on the CTRS for other age groups, as well as validating the CTRS-15 on another sample of preschool children and cross-validating this measure with other types of inattention measures.

A final area of interest raised by this study is the stability of the relation between inattention and academic problems in children. A substantial body of research has indicated that high levels of inattention are related to lower academic performance (Frick, et al., 1991; Lonigan, et al., 1999; Willcutt & Pennington, 2000). However, the overlap between items that directly measure inattention and academic performance on behavioral rating scales may have created an artificial relation due to lack of discriminant validity of these measures. The relation between these two domains should be reexamined utilizing multiple measures of inattention, including the CTRS-15, to delineate between the direct effects, if any, of inattention on academic performance and the artificial effects resulting from inefficient measurement.

Conclusion

In conclusion, the results of this study indicate that the CTRS can be reduced in length with an enhanced ability to discriminate effectively across the behavioral spectrum. The CTRS-15 may, in fact, be more accurate than other versions of the CTRS because items that were only marginally informative or inappropriate for use with preschool children were removed, reducing potential measurement error. Further, the practical benefits derived from the CTRS-15, such as the reduced length of the measure, may increase the use of the measure in research conducted in classrooms as well as potentially increase teacher compliance in such research. Therefore, the use of the CTRS-15 as a behavioral measure in research studies within classrooms may add valuable knowledge to the fields of behavioral and education research, as well as contribute to the investigation of the relation between these two areas.

Acknowledgments

This work was supported by grants from the National Institute of Child Health and Human Development (HD38880, HD30988) and the Institute of Education Science, U.S. Department of Education (R305B04074). Views expressed herein are solely those of the authors and have not been reviewed or cleared by the grantors. We thank Christopher Schatschneider for his helpful comments and guidance.

Appendix

Conners Teacher Rating Scale - 15

	Not True at All Never, Seldom	Just a Little True Occasionally	Pretty Much True Often, Quite a Bit	Very Much True Very Often, Very Frequent
1. Restless in the "squirmy" sense	0	1	2	3
2. Restless, always up and on the go	0	1	2	3
3. Cannot remain still	0	1	2	3
4. Runs about or climbs excessively in situations where it is inappropriate	0	1	2	3
5. Has difficulty playing or engaging in leisure activities quietly	0	1	2	3
6. Fails to finish things that s/he starts	0	1	2	3
7. Inattentive, easily distracted	0	1	2	3
8. Forgets things she/he has already learned	0	1	2	3
9. Short attention span	0	1	2	3
10. Does not follow through on instructions and fails to	0	1	2	3

	Not True at All Never, Seldom	Just a Little True Occasionally	Pretty Much True Often, Quite a Bit	Very Much True Very Often, Very Frequent
finish schoolwork (not due to oppositional behavior or failure to understand instructions)				
11. Pouts and sulks	0	1	2	3
12. Mood changes quickly	0	1	2	3
13. Uncooperative	0	1	2	3
14. Defiant	0	1	2	3
15. Spiteful or vindictive	0	1	2	3

References

- Administration on Children, Youth and Families. First Progress Report of the Head Start Program Performance Measures. U.S. DHHS; Washington, D.C.: 1997. Head Start Bureau.
- American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. Fourth Edition, Text Revision. American Psychiatric Association; Washington, DC: 2000.
- Anastopoulos AD, Shelton TL, DuPaul GJ, Guevremont DC. Parent training for attention-deficit hyperactivity disorder: Its impact on parent functioning. *Journal of Abnormal Child Psychology*. 1993; 21:581–596. [PubMed: 8294653]
- Conners, CK. Conners' Rating Scales – Revised: Long Form. Multi-Heath Systems; North Tonawanda, NY: 1997.
- Conners, CK. Conners' Rating Scales – Revised: Short Form. Multi-Heath Systems; North Tonawanda, NY: 1997b.
- Conners CK. A teacher rating scale for use in drug studies with children. *American Journal of Psychiatry*. 1969; 126:884–888. [PubMed: 4900822]
- Cuffe SP, Moore CG, McKeown RE. Prevalence and correlates of ADHD symptoms in the National Health Interview Survey. *Journal of Attention Disorders*. 2005; 9:392–401. [PubMed: 16371662]
- Darling-Hammond, L. Solving the Dilemmas of Teacher Supply, Demand, and Standards: How We Can Ensure a Competent, Caring, and Qualified Teacher for Every Child. National Commission on Teacher & America's Future; New York: 2000.
- Frick PJ, Kamphaus RW, Lahey BB, Loeber R, Christ MAG, Hart EL, et al. Academic underachievement and the disruptive behavior disorders. *Journal of Consulting and Clinical Psychology*. 1991; 59:289–294. [PubMed: 2030190]
- Gerhardstein RR, Lonigan JL, Cukrowicz KC, McGuffey JA. Factor structure of the Conners' Teacher Rating Scale – Short Form in a low-income preschool sample. *Journal of Psychoeducational Assessment*. 2003; 21:223–243.
- Goyette CH, Conners CK, Ulrich RF. Normative data on revised Conners parent and teacher rating scales. *Journal of Abnormal Child Psychology*. 1978; 6:221–236. [PubMed: 670589]
- Hambleton, RK.; Swaminathan, H.; Rogers, HJ. Fundamentals of Item Response Theory. Sage Publications; Newbury Park: 1981.
- Hinshaw SP. Externalizing behavior problems and academic underachievement in childhood and adolescence: Causal relationships and underlying mechanisms. *Psychological Bulletin*. 1992; 111:127–155. [PubMed: 1539086]
- Lahey BB, Pelham WE, Loney J, Lee SS, Willcutt E. Instability of the DSM-IV Subtypes of ADHD From Preschool Through Elementary School. *Archives of General Psychiatry*. 2005; 62:896–902. [PubMed: 16061767]

- Lahey BB, Pelham WE, Loney J, Kipp H, Ehrhardt A, Steve S, et al. Three-year predictive validity of DSM-IV attention deficit hyperactivity disorder in children diagnosed at 4-6 years of age. *American Journal of Psychiatry*. 2004; 161:2014–2020. [PubMed: 15514401]
- Lonigan CJ, Bloomfield BG, Anthony JL, Bacon KD, Phillips BM, Samwel CS. Relations among emergent literacy skills, behavior problems, and social competencies in preschool children from low- and middle-income backgrounds. *Topics in Early Childhood Special Education*. 1999; 19:40–54.
- Lord FM. Practical applications of item characteristic curve theory. *Journal of Educational Measurement*. 1977; 14:117–138.
- Lord FM, Novick MR. Standard errors of measurement at different ability levels. *Journal of Educational Measurement*. 1968a; 21:239–243.
- Lord, FM.; Novick, MR. *Statistical theories of mental test scores*. Addison-Wesley; Reading, MA: 1968b.
- Muthén, LK.; Muthén, BO. *Mplus 3.12*. Muthén & Muthén; Los Angeles, CA: 2005. [Computer program]
- NICHHD Early Child Care Research Network. The relation of child care to cognitive and language development. *Child Development*. 2000; 71:960–980. [PubMed: 11016559]
- Parker, Sitarenios, Conners. Abbreviated Conners' Rating Scales revisited: A confirmatory factor analytic study. *Journal of Attention Disorders*. 1996; 1:55–62.
- Pelham WE, Milich R, Murphy DA, Murphy HA. Normative data on the IOWA Conners Teacher Rating Scale. *Journal of Clinical Child Psychology*. 1989; 18:259–262.
- Samejima, F. Estimation of latent ability using a response pattern of graded scores. *Psychometric Society; Iowa City, IA: 1969. Psychometric Monograph No. 17*
- Schafer, JL. *NORM*. Version 2.03. Pennsylvania State University; University Park, PA: 2000.
- Spira EG, Fischel JE. The impact of preschool inattention, hyperactivity, and impulsivity on social and academic development: A review. *Journal of Child Psychology and Psychiatry*. 2005; 46(7):755–773. [PubMed: 15972069]
- Sprague RL, Sleator EK. Effects of psychopharmacological agents on learning disorders. *Pediatric Clinics of North America*. 1973; 20:719–735. [PubMed: 4582121]
- Stark, S. University of Illinois; Urbana-Champaign: 2001. *MODFIT: A computer program for model-data fit*. Unpublished manuscript
- Whitehurst GJ, Falco FL, Lonigan CJ, Fischel JE, DeBaryshe BD, Valdez-Menchaca MC, et al. Accelerating language development through picture book reading. *Developmental Psychology*. 1988; 24:552–559.
- Willcutt EG, Pennington BF. Comorbidity of reading disability and attention-deficit/hyperactivity disorder: Differences by gender and subtype. *Journal of Learning Disabilities*. 2000:179–191. [PubMed: 15505947]
- Willoughby MT. Developmental course of ADHD symptomology during the transition from childhood to adolescence: A review with recommendations. *Journal of Child Psychology and Psychiatry*. 2003; 44:88–106. [PubMed: 12553414]

Figure 1.

In panel 1, the Test Information Functions for full behavior scales show the amount of information provided by each behavior scale across the behavioral spectrum. Each behavior scale, prior to removal of poor fitting and extraneous items, appears to be close to a normal curve except each scale provides a varying amount of total information. In panel 2, the Target Information Function and Test Information Functions of Revised Scales are presented. Each behavior scale, after being reduced to five items each, appears to approximate the normal curve and provides comparable information to the other two scales as well as the Target Test Information Function.

Table 1

Discrimination parameters (“a” parameters) and Difficulty parameters (“b” parameters) for each item within each behavior factor.

Item	<i>a</i>	SE	<i>b1</i>
<i>Inattention</i>			
Item 9 – Daydreams	0.634	0.031	0.423
Item 16 - Appears to be easily led by other children	0.682	0.025	−0.314
Item 17 - Appears to lack leadership	0.705	0.025	0.060
Item 18 * - Fails to finish things that s/he starts	0.849	0.015	0.261
Item 21 - Easily frustrated in efforts	0.705	0.026	0.233
Item 23 - Difficulty in learning	0.846	0.018	0.649
Item 24 * - Inattentive, easily distracted	0.905	0.012	−0.109
Item 26 * - Forgets things she/he has already learned	0.803	0.022	0.672
Item 35 * - Short attention span	0.890	0.013	0.002
Item 39 - Lacks interest in schoolwork	0.773	0.024	0.766
Item 44 * - Has difficulty engaging in leisure activities quietly	0.858	0.018	0.528
<i>Hyperactivity/Impulsivity</i>			
Item 1 * - Restless in the “squirmy” sense	0.888	0.010	−0.356
Item 2 - Makes inappropriate noises when s/he shouldn’t	0.864	0.011	0.228
Item 3 - Demands must be met immediately	0.786	0.013	−0.028
Item 7 - Distractibility or attention span a problem	0.847	0.014	−0.138
Item 8 - Disturbs other children	0.870	0.026	0.030
Item 13 * - Restless, always up and on the go	0.924	0.019	0.122
Item 14 - Excitable, impulsive	0.875	0.011	0.074
Item 15 - Excessive demands for attention	0.812	0.021	0.339
Item 19 - Childish and immature	0.758	0.020	0.008
Item 20 - Denies mistakes or blames others	0.685	0.022	0.032
Item 28 - Is always “on the go” or acts as if driven by a motor	0.903	0.013	0.350
Item 30 * - Cannot remain still	0.929	0.019	−0.057
Item 32 - Leaves seat when remaining seated is expected	0.812	0.016	0.536
Item 33 - Fidgets with hands or feet or squirms in seat	0.886	0.009	−0.038
Item 37 - Only pays attention to things he/she is really interested in	0.782	0.008	0.119
Item 38 - Has difficulty waiting his/her turn	0.814	0.015	−0.065
Item 40 * - Runs about or climbs excessively	0.889	0.017	0.504
Item 42 - Interrupts or intrudes on others	0.821	0.013	−0.061
Item 43 * - Difficulty playing or engaging in leisure activities quietly	0.827	0.018	0.342
<i>Oppositional Defiant Behaviors</i>			
Item 4 - Acts “smart” (impudent or sassy)	0.833	0.034	0.335
Item 5 - Temper outbursts and unpredictable behavior	0.882	0.008	0.542
Item 6 - Overly sensitive to criticism	0.536	0.018	−0.082
Item 10 * - Pouts and sulks	0.822	0.015	0.045

Item	<i>a</i>	SE	b1
Item 11 * - Mood changes quickly	0.873	0.018	0.408
Item 12 – Quarrelsome	0.898	0.015	0.447
Item 22 * - Uncooperative	0.928	0.008	0.487
Item 25 * - Defiant	0.946	0.013	0.433
Item 27 - Actively defies or refuses to comply with adults' requests	0.942	0.009	0.476
Item 31 * - Spiteful or vindictive	0.841	0.015	0.864
Item 36 - Argues with adults	0.870	0.021	0.564

* *Note.* Indicates item was selected for the CTRS-15.

Table 2

Correlations between rating scales within each behavioral domain.

Behavioral Domain/Scale	Behavioral Domain/Scale									
	1	2	3	4	5	6	7	8	9	10
Inattention										
1. CTRS-15	—	—	—	—	—	—	—	—	—	—
2. CTRS-R:S	.32*	—	—	—	—	—	—	—	—	—
3. Hybrid CTRS	.92*	.34*	—	—	—	—	—	—	—	—
4. ADHD-IV Rating Scale Hyperactivity/Impulsivity	.18*	.06	.20*	—	—	—	—	—	—	—
5. CTRS-15	—	—	—	—	—	—	—	—	—	—
6. CTRS-R:S	.94*	—	—	—	—	—	—	—	—	—
7. Hybrid CTRS	.94*	.96*	—	—	—	—	—	—	—	—
8. ADHD-IV Rating Scale Opposition	.20*	.22*	.24*	—	—	—	—	—	—	—
9. CTRS-15	—	—	—	—	—	—	—	—	—	—
10. CTRS-R:S	.91*	—	—	—	—	—	—	—	—	—
11. Hybrid CTRS	.96*	.94*	—	—	—	—	—	—	—	—

* Indicates correlation is significant at $p < .001$.