Published in final edited form as:
Psychol Assess. 2011 September ; 23(3): 778–791. doi:10.1037/a0023290.

Colorado Learning Difficulties Questionnaire:Validation of a parent-report screening measure

Erik G. Willcutt
Department of Psychology and Neuroscience, University of Colorado, Boulder

Richard Boada
Department of Neurology, University of Colorado at Denver and Health Sciences Center, and Department of Psychology, University of Denver

Margaret W. Riddle
Department of Psychology, University of Denver

Nomita Chhabildas and **John C. DeFries**
Department of Psychology and Neuroscience, University of Colorado, Boulder

Bruce F. Pennington
University of Denver

Abstract

This study evaluated the internal structure and convergent and discriminant evidence for the Colorado Learning Difficulties Questionnaire (CLDQ), a 20-item parent-report rating scale that was developed to provide a brief screening measure for learning difficulties. CLDQ ratings were obtained from parents of children in two large community samples and two samples from clinics that specialize in the assessment of learning disabilities and related disorders (total N = 8,004). Exploratory and confirmatory factor analyses revealed five correlated but separable dimensions that were labeled *reading*, *math*, *social cognition*, *social anxiety*, and *spatial difficulties*. Results revealed strong convergent and discriminant evidence for the CLDQ Reading scale, suggesting that this scale may provide a useful method to screen for reading difficulties in both research studies and clinical settings. Results are also promising for the other four CLDQ scales, but additional research is needed to refine each of these measures.

Keywords

reading; math; learning; rating scale; screening

Learning disorders (LDs) are defined by significant academic underachievement that is unexpected based on an individual's age, cognitive ability, and education (e.g., American Psychiatric Association, 2000). The fourth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV; American Psychiatric Association, 2000) provides diagnostic criteria for Reading Disorder (RD), Math Disorder (MD), and Disorder of

Point of Contact: Correspondence concerning this article should be addressed to Erik Willcutt, Department of Psychology and Neuroscience, UCB 345, University of Colorado, Boulder, 80309. willcutt@colorado.edu..

Written Expression. In addition to these DSM-IV categories, other authors described non-verbal learning disability (NVLD), a syndrome characterized by specific difficulties in mathematics and spatial functioning, along with impairments in social cognition similar to the difficulties exhibited by individuals with pervasive developmental disorders (PDD; e.g., Klin, Volkmar, Sparrow, Cicchetti, & Rourke, 1995; Rourke, 1989).

LDs are associated with a range of negative outcomes and significant publich health costs. Prevalence estimates suggest that 5–15% of the population meet criteria for at least one LD (e.g., American Psychiatric Association, 2000; Gross-Tsur, Manor, & Shalev, 1996; Rutter et al., 2004; Shaywitz, Shaywitz, Fletcher, & Escobar, 1990), and over half of all students who receive special education services are identified due to an LD (e.g., Schnoes, Reid, Wagner, & Marder, 2006). Studies that compared groups with and without an LD found that individuals with an LD experience greater academic difficulties, report lower motivation and greater frustration and distress in school, are more likely to drop out of high school prior to graduation, and reach lower levels of educational and occupational attainment as adults (e.g., Boetsch, Green, & Pennington, 1996; Daniel et al., 2006; Goldston et al., 2007; McGee, Prior, Willams, Smart, & Sanson, 2002; Willcutt et al., 2007). LDs also co-occur more often than expected by chance with one another and with other disorders such as attention-deficit/hyperactivity disorder (ADHD), conduct disorder, anxiety disorders, and depression (Antshel & Khan, 2008; Daniel et al., 2006; Maughan, Rowe, Loeber, & Stouthamer-Loeber, 2003; McGee et al., 2002; Semrud-Clikeman et al., 1992; Trzesniewski, Moffitt, Caspi, Taylor, & Maughan, 2006; Willcutt et al., 2007; Willcutt & Pennington, 2000a; Willcutt & Pennington, 2000b).

The high prevalence of LDs and their frequent co-occurrence with other disorders suggests that LD assessment measures should be systematically included in clinical assessment batteries and research studies focusing on developmental disorders. However, a full LD assessment requires the administration of standardized tests of academic achievement and cognitive ability by a trained examiner in a one-on-one testing session that typically lasts several hours. It is not feasible to complete such an extensive evaluation as part of many clinical assessments and research studies, particularly if comorbid learning difficulties are not the primary referral question for a clinical assessment or are a secondary aim of a study focusing on a related but separate topic.

Similar challenges are faced by clinicians or researchers who wish to screen systematically for a range of psychopathology as part of a standard clinical assessment battery or research protocol, as it is often unrealistic to devote the time necessary to obtain a comprehensive assessment of all relevant disorders. To address this issue, several screening measures for developmental psychopathology have been developed, such as the Achenbach System of Empirically Based Assessment (ASEBA; Achenbach & Rescorla, 2001), the Behavior Assessment System for Children (BASC; Reynolds & Kamphaus, 2004), the Conners Rating Scales (e.g., Conners, Sitarenios, Parker, & Epstein, 1998) and the Early Childhood Inventory (ECI), Child Symptom Inventory (CSI), and Adolescent Symptom Inventory (ASI) developed by Gadow and colleagues (e.g., Gadow & Sprafkin, 1997a; Gadow & Sprafkin, 1997b; Gadow & Sprafkin, 1998). Each of these measures can be completed quickly by parents or teachers to screen efficiently for a broad range of psychopathology, and all are used widely in both research studies and clinical practice. Scores from these measures do not replace diagnostic interviews, and are not intended to provide clinical diagnoses or to guide treatment planning in isolation. Instead, these norm-referenced rating scales provide reliable and valid indicators of areas in which an individual appears to be experiencing significant difficulty in comparison to others the same age, and these areas can then be targeted directly for more intensive evaluation.

In contrast to these well-validated screening measures for psychopathology, to our knowledge there are no scales designed to screen for specific learning disorders and related developmental difficulties. In this manuscript we describe the development of the Colorado Learning Difficulties Questionnaire (CLDQ), a parent-report rating scale that may provide a useful screening instrument for use in clinical settings and research studies. The CLDQ was designed to assess specific dimensions of functioning that are most often impaired in children with learning difficulties, including reading, math, social cognition, spatial functioning, and memory. Data from four large samples (total N =8,004) were used to evaluate the internal structure and convergent and discriminant evidence for the CLDQ scales. Specific goals were as follows:

1. To assess the number of dimensions of learning difficulties assessed by the CLDQ, initial exploratory factor analyses (EFA) were completed in each sample, and a subsequent multigroup confirmatory factor analysis was used to test whether the factor structure could be equated across the four samples. We hypothesized that these analyses would identify separable dimensions of reading, math, spatial functioning, social cognition, and memory.
2. The inter-rater reliability of each CLDQ scale was evaluated by examining correlations between maternal and paternal ratings, and estimates of test-retest reliability were obtained from maternal ratings completed approximately one year apart.
3. The four datasets included a range of external measures of each of the constructs assessed by the CLDQ. Convergent and discriminant evidence for the CLDQ scales was evaluated by testing whether correlations were significantly higher between each scale and external measures of the same construct than measures of other constructs.
4. CLDQ scores of groups with RD, MD, NVLD, and other developmental disorders were compared to test whether predicted associations were observed between each CLDQ scale and specific disorders. We expected that individuals with RD would exhibit higher elevations on a CLDQ Reading scale than any other scale that emerged in factor analyses of the CLDQ, whereas individuals with MD would exhibit the most pronounced elevations on a Math scale. Groups with NVLD were expected to score highest on on a CLDQ scale measuring spatial difficulties, and groups with NVLD or a pervasive developmental disorder (PDD) were expected to exhibit the most pronounced impairment on a CLDQ scale measuring social cognition.

Method

Participants

Parents of children and adolescents in two clinic samples and two community samples completed the CLDQ as part of a larger packet of questionnaires. Descriptive characteristics of the samples are summarized in Table 1.

University of Denver Developmental Neuropsychology Clinic—This sample includes 954 consecutive referrals to a University clinic specializing in neuropsychological assessments of children and adolescents. Although the most frequent referral questions are RD and ADHD, the sample included cases with a range of developmental disorders (Table 1).

University of Colorado at Boulder Attention, Behavior, and Learning Clinic— This second clinic-referred sample includes 179 consecutive referrals between 6 and 18 years old. The Boulder clinic specializes in the assessment of ADHD and learning disabilities, but also sees cases with a range of referral concerns (Table 1).

Twin sample—Parents completed the CLDQ as part of the Colorado Learning Disabilities Research Center twin study (CLDRC), an ongoing study of the etiology of learning and attentional difficulties (e.g., DeFries et al., 1997; Willcutt, Pennington, Olson, Chhabildas, & Hulslander, 2005). Based on an initial screening of over 4,000 twin pairs, pairs between 8 and 18 years old were recruited if at least one of the twins met criteria for reading disability or DSM-IV ADHD (N = 972), and a matched comparison sample of twin pairs without RD or ADHD was recruited from the same schools (N = 868; see Willcutt et al., 2005 for a full description of the recruitment procedures). Each member of the pair then completed a detailed assessment battery that included measures of general cognitive ability, reading and math achievement, social functioning, and internalizing and externalizing psychopathology. Because more mothers (96%) than fathers (78%) completed the CLDQ, maternal ratings were used for all analyses except tests of inter-rater reliability, which examined the correlation between ratings by the two parents.

Community screening sample—As part of a larger study of the DSM-IV ADHD subtypes, parents of all children attending schools in five local public school districts were invited to participate in the first phase of the study by completing an initial screening questionnaire that included the CLDQ (N = 5,031 completed the questionnaire). A subset of families of children with and without DSM-IV ADHD were then invited to participate in a more extensive individual testing session that included the measures of intelligence and academic achievement that were used to evaluate the convergent and discriminant evidence for the CLDQ scores. The individual assessment was completed by 502 participants with ADHD, 532 of their biological siblings, and a comparison sample of 530 children without ADHD matched to the ADHD sample on age, sex, ethnicity, socioeconomic status, and school.

Inclusion criteria—In addition to the inclusion criteria applied as part of each individual study, several additional criteria were required for a case to be included in the current analyses. In the clinic samples, the CLDQ was typically not administered to parents of individuals older than 18 years of age, and most parents of children younger than 6 years old were unable to answer several items that were not yet developmentally typical (e.g., difficulty with spelling or handwriting). Therefore, analyses of the clinic samples were restricted to individuals between 6 and 18 years old. In all samples a small subset of parents failed to complete three or more of the items on the final 20-item CLDQ scale (0.1 – 0.6% of all questionnaires across studies). In addition, two parents of children in the Denver clinic sample (0.2%), two parents from the twin study (0.1%), and four parents of children in the community screening sample (0.1%) circled multiple answers for several CLDQ items. These cases (0.1 – 0.7% of all individuals) were excluded from all analyses and are not included in the samples described in Table 1.

Development of the CLDQ

The initial item pool—The CLDQ was initially developed to quantify the presenting concerns of parents when they brought their child for a psychoeducational or neuropsychological evaluation at the Denver clinic. The scale was included as part of a developmental and family history questionnaire completed by all parents at the beginning of each assessment. Items on the initial CLDQ were designed to assess functioning in eight domains: reading, math, attention / hyperactivity, anxiety, depression, social functioning,

spatial ability, and memory. Parents answered each question on a five-point Likert scale with the following anchors: (1) *never / not at all*, (2) *rarely / a little*, (3) *sometimes*, (4) *frequently / quite a lot*, and (5) *always / a great deal*.

The 46 items on the initial questionnaire are listed in Table 2. Parents of children in the Denver clinic sample and the twin sample completed the full 46-item scale. Over half of these items were dropped for theoretical reasons or due to weak psychometric characteristics, leaving a final 20-item scale that was completed by parents of the community sample and Boulder clinic sample. In the remainder of this section we briefly describe the rationale for the exclusion of the other 26 questions from the original pool of items.

Elimination of ADHD and depression items—At the time the scale was developed, standardized measures were not available to screen for symptoms of ADHD or depression. Therefore, the original scale included 11 items that assessed behaviors related to ADHD and 5 items designed to assess depression (items 21 – 36 in Table 2). Since that time, more comprehensive ADHD and depression screening instruments have been published (e.g., Barkley & Murphy, 1998;DuPaul, Power, Anastopoulos, & Reid, 1998;Gadow & Sprafkin, 1997a;Kovacs, 1988), and preliminary analyses of the CLDQ indicated that the psychometric properties of ADHD and depression composites based on CLDQ items were weaker than the characteristics of the existing scales. Therefore, the 16 ADHD and depression items from the original CLDQ were dropped from the current version of the scale. Initial factor analyses including these items indicated that all ADHD and depression items on the CLDQ loaded on factors separate from the final factors described in this report, and the overall factor structure of the remaining items remained the same whether or not the ADHD and depression items were included in the analysis.

Exclusion of additional items—After removing the items designed to assess ADHD and depression, the psychometric characteristics of the remaining items were examined carefully. Nine items did not load on any of the factors in initial factor analyses (all loadings < .40), and several of these items also had low inter-rater and test-retest reliability (items 37 – 46 in table 2). Therefore, these items were also dropped from the final version of the scale described in this paper.

Internal structure of the final scale—An initial exploratory factor analysis (EFA) was conducted separately in each sample. Principal axis extraction and direct oblimin rotation were used to extract factors with eigenvalues greater than one. The direct oblimin rotation was used because it is an oblique rotation that permits the obtained factors to correlate, and therefore requires fewer a priori assumptions about the relations among the variables than an orthogonal method of rotation. However, the same number of factors and similar factor loadings were obtained when a principal components analysis with varimax rotation was conducted, suggesting that the results are robust across different methods of factor extraction and rotation.

A five-factor solution best explained the data in all four samples (Table 3). The factors were labeled *Reading*, *Math*, *Social Cognition*, *Social Anxiety*, and *Spatial.* All 20 items loaded highest on their primary factor in all samples, and only the item assessing friendship difficulties cross-loaded on any other factor (it loaded on both the Social Anxiety and Social Cognition factors in the Denver clinic sample and the twin sample).

After conducting the EFAs to obtain an initial appraisal of the structure of the CLDQ in each sample, a confirmatory factor analysis (CFA) model was fitted to test directly whether the factor structure could be equated across the four samples. The item loadings and factor

covariances were constrained to be equal in all samples, whereas the means and variances of the CLDQ items were not equated because these parameters were expected to differ in community and clinic samples. Due to the large samples included in the analysis, the fit of the CFA model was evaluated with the comparative fit index (CFI; Bentler, 1990) and root mean square error of approximation (RMSEA; Browne & Cudeck, 1993), two fit indices that are less sensitive to sample size than other fit indices such as $\times^2$ (e.g., Fan, Thompson, & Wang, 1999). Although cutoffs used to assess goodness-of-fit are based primarily on convention (e.g., Chen, Curran, Bollen, Kirby, & Paxton, 2008), widely-used thresholds for good model fit are RMSEA less than or equal to .05 and CFI greater than .90 (e.g., Schumacker & Lomax, 2004). The fit of the constrained model was adequate based on both of these indices (*CFI* = .931; *RMSEA* = .042, 95% CI [.041, .043]), and only slightly worse than the fit of a model in which the item loadings and factor covariances were unconstrained (*CFI* = .940; *RMSEA* = .041), providing additional support for the hypothesis that the internal structure of the CLDQ is similar in the four samples.

Reliability of the CLDQ scales—Based on the results of the EFA and CFA, five CLDQ scale scores were calculated by computing the mean of the items that loaded on each factor. Inter-rater reliability was assessed by the correlation between mother and father ratings in the twin sample, and test-retest reliability was assessed over an interval of approximately one year in a subset of the community sample who returned for a follow-up assessment as part of the larger study (N = 554). Estimates of internal consistency and reliability were high for the Reading scale items and composite score and moderate for the other four scales (Tables 3 and 4).

Measures to evaluate the convergent and discriminant evidence of the CLDQ scores

Because none of the study protocols were designed to evaluate the CLDQ, the specific measures available to evaluate the convergent and discriminant evidence for each CLDQ scale varied across samples. Nonetheless, each of the samples included at least one measure relevant to each of the five CLDQ domains, and most samples included two or more measures of each construct (Table 5). Due to space constraints it is not possible to describe all of these external measures in detail. Therefore, in the remainder of this section we briefly describe each test or scale, and provide references for additional information about the measures in the notes for Table 5.

Reading and math achievement—Measures of single-word reading, reading comprehension, math calculations, and math word problems were obtained from the Woodcock-Johnson Tests of Achievement, the Peabody Individual Achievement Test, the Gray Oral Reading Test, and the Wide Range Achievement Test, all of which are widely-used standardized measures of academic achievement.

Social functioning—The Behavior Assessment System for Children (BASC) and Achenbach System of Empirically Based Assessment (ASEBA) are nationally-normed parent and teacher rating scales that include measures of social functioning. The sociometric rating scale developed by Dishion (1990) asks the child's teacher to estimate the proportion of students in the class who like, dislike, or ignore the child.

Spatial functioning—All four studies administered the Block Design subtest from one of the Wechsler Intelligence Scales, and participants in three of the four samples also completed the Rey-Osterreith Complex Figure Test (ROCFT), a task which requires the participant to copy a complex figure. A subset of the Denver clinic sample completed the Developmental Test of Visual-Motor Integration (DTVMI), a standardized measure that requires the participant to copy a series of increasingly complex designs. Finally, the twin

study included Primary Mental Abilities (PMA) Spatial Relations subtest, a test that requires the participant to select from five choices the figure that is a clockwise rotation of a target figure.

Psychopathology—Measures of several dimensions of psychopathology that frequently co-occur with learning difficulties were analyzed to further evaluate the discriminant evidence for the CLDQ scales. Although different measures were used to assess ADHD in the four samples, each of these measures provides composite scores derived from parent and teacher ratings of *DSM-IV* inattention and hyperactivity-impulsivity symptoms. Parents and teachers also completed the internalizing and externalizing scales on the ASEBA or BASC, and parents rated symptoms of generalized anxiety disorder (GAD), separation anxiety disorder (SAD), major depressive disorder (MDD), and pervasive developmental disorder on the Adolescent Symptom Inventory (ASI), Child Symptom Inventory (CSI), or Diagnostic Interview for Children and Adolescents (DICA-IV).

Data preparation and consolidation

Data Adjustments—The distribution of each variable was assessed for significant deviation from normality, and an appropriate transformation was applied to approximate a normal distribution for variables with skewness or kurtosis greater than one. No scores on the CLDQ or the measures used for external validation met our a priori criteria for outlying values (more than three SD below the mean and more than 0.5 SD beyond the next most extreme score).

There were small but significant correlations between age and the CLDQ Reading scale, $r = -.08$; 95% CI [−.11, −.06], Math scale, $r = -.09$; 95% CI [−.14, −.04], and Spatial scale, $r = -.07$; 95% CI [−.11, −.03], and several of the external measures (r =.06–.13). Therefore, an age-adjusted score was created for each measure by regressing the variable onto age and computing the residual score. To test further for potential differences in results as a function of age, primary analyses were also conducted separately in subsets of each sample divided by age (younger than 11 years old, 11 – 13 years old, and older than 13 years old). Although some of these analyses were constrained by small sample sizes, the pattern of results was extremely similar in all age groups. Therefore, results are reported for the full samples (results for the separate age groups are available from the lead author).

Creation of composites for the constructs used for external validation—Because initial analyses revealed that the pattern of results was nearly always similar when multiple measures of an external construct were analyzed separately, composite scores were created for several of the constructs that were assessed by multiple measures. Each composite score is the mean of age-regressed standardized scores on all measures of the construct that were administered in a particular sample. The *reading composite* is the mean of the measures of single-word reading and reading comprehension, and the *math composite* is the mean of the measures of math calculations and word problems. The *social isolation composite* includes the ratings of withdrawn behavior and the extent to which the individual is ignored by peers, the *social rejection composite* is the mean of the Social Problems scale and the teacher rating of the proportion of peers who dislike the participant, and the *social strengths composite* is the mean of the measures of social skills and teacher ratings of the proportion of peers who like the individual. The *anxiety composite* is the mean of the ASEBA / BASC scale and parent ratings of GAD and SAD. The *spatial composite* includes Block Design, the copy trial from the ROCFT, the DTVMI, and PMA Spatial Relations.

Data analyses

Corrections for non-independence in the twin sample—Phenotypic analyses of twin data must account for the fact that the two twins in a pair are not completely independent. Therefore, a multilevel approach was used that considered nesting of twins within families (Muthen & Muthen, 2009) to provide valid estimates of population parameters, measures of association between variables, and tests of significance.

Meta-analytic procedures to calculate overall effect sizes across samples—To simplify interpretation and minimize the number of statistical tests, meta-analytic procedures were used to compute a single summary statistic and confidence interval to describe the relation between each CLDQ scale and the composite measure of each external construct. In the first step of this procedure a separate effect size is calculated for each sample (*r* for correlational analyses of continuous measures and Cohen's *d* (1988) for comparisons between means of the clinical groups). If the effect sizes in the four samples are homogenous, an overall effect can be calculated using a fixed effects model that weights each individual effect size by the corresponding sample size (e.g., Hedges & Olkin, 1985). If there is significant heterogeneity among the samples, however, the confidence interval obtained from the fixed effects model may be underestimated (e.g., Higgins & Thompson, 2002).

We tested for significant heterogeneity among the samples by calculating *Q*, an estimate of the variability of individual effect sizes around the overall estimated effect size (DerSimonian & Laird, 1986). Although *Q* was not significant for most analyses, significant heterogeneity ($P < .05$) was observed for three effects (correlations between both inattention and hyperactivity-impulsivity and CLDQ Social Cognition, along with mean differences between groups with and without RD on the CLDQ Reading scale), and estimates of heterogeneity approached significance in several additional analyses ($P < .10$). Therefore, the random effects model described by DerSimonian and Laird (1986) was used to estimate each overall effect size and corresponding confidence interval (r_w for dimensional analyses and d_w for comparisons of group means). The random effects model is a more conservative approach that adjusts for heterogeneity by weighting each effect size by both the inverse variance of that sample and an additional weight based on *Q*. If *Q* is low the additional weight becomes zero, and the fixed effects and random effects models yield identical results.

Analytic plan—The EFA and CFA described previously support the internal structure of the CLDQ scales. Convergent evidence for each CLDQ scale was first evaluated by testing if scores on the scale were significantly correlated with independent measures of the same theoretical construct (for example, if the CLDQ reading scale was correlated with performance on standardized measures of reading achievement). In addition, CLDQ scores in the clinical groups were compared to the population mean estimated from the community screening sample to test if scores on each CLDQ scale were significantly elevated in groups that are known to have a specific weakness in that domain of functioning (e.g., math scores in groups with MD). Discriminant evidence for the scales was then evaluated by testing for the predicted differential associations between the CLDQ scales and the external measures and clinical disorders.

Results

Scores on all five CLDQ scales were significantly correlated with nearly all external measures (Table 6), and ratings of all clinical groups were significantly higher than the estimated population mean on all CLDQ scales (Table 7). These results clearly indicate that the CLDQ is sensitive to clinical status, providing preliminary convergent evidence for the

CLDQ scales. On the other hand, the ubiquitous associations between all CLDQ scales and all external measures and clinical diagnoses underscore the need to examine carefully the discriminant evidence for each CLDQ scale.

Reading scale

The CLDQ Reading scale was highly correlated with composite measures of reading achievement in all four samples (r_w = .64), and nonoverlapping confidence intervals indicated that these correlations were significantly higher than the correlations between the Reading scale and all other domains of functioning (Table 6). Similarly, the effect size of the difference between the RD group and the estimated population mean was large (d_w = 1.81; Table 7), substantially higher than the moderate effect sizes obtained for the RD group on the other CLDQ scales (d_w = .31 −.82), and significantly higher than the means of groups with other disorders. These results provide strong convergent and discriminant support for the CLDQ Reading scale.

Math scale

Measures of math achievement were more highly correlated with the CLDQ Math Scale than the other four CLDQ scales (Table 6), although the magnitudes of these correlations are lower than the correlations between the CLDQ Reading scale and the reading achievement composites. Similarly, groups with MD or NVLD scored significantly higher on the Math scale than any other clinical group (Table 7), and in the group with MD the effect size on the Math scale was significantly larger than the effect on any other CLDQ scale.

Social Cognition scale

As predicted, the Social Cognition scale was more highly correlated with weak social skills, social rejection, and symptoms of PDD than any other CLDQ scale, but correlations were also unexpectedly high between the Social Cognition scale and measures of externalizing symptoms (Table 6). Group comparisons indicated that groups with a PDD scored significantly higher on the Social Cognition scale than any of the other clinical groups, and were more impaired on the scale than any of the other CLDQ scales (Table 7).

Social Anxiety scale

Because this scale unexpectedly separated from the Social Cognition scale in the factor analysis, the discriminant evidence for these scales was carefully examined. The CLDQ Social Anxiety scale was most strongly associated with parent and teacher ratings of anxiety and social isolation (Tables 6), providing convergent evidence for the Social Anxiety scale. In contrast to the stronger associations between the Social Cognition scale and symptoms of PDD and externalizing disortders, CLDQ Social Anxiety scores were more strongly associated with social isolation, withdrawn behaviors, and anxiety disorders (Tables 6 and 7).

Spatial scale

The CLDQ Spatial scale was more highly correlated with the external measures of spatial functioning than the Reading, Social Cognition, or Social Anxiety scales (Table 6), but this association was similar in magnitude to the correlation between the CLDQ Math Scale and the spatial composite. Further, the correlation between the CLDQ Spatial scale and the external measures of spatial functioning was significantly lower than the correlation between the Spatial scale and inattention, and was similar to the correlations between the Spatial scale and measures of hyperactivity-impulsivity symptoms and math achievement. The strongest discriminant evidence for the CLDQ Spatial scale was provided by the large effect size in the group with NVLD (Table 7). However, consistent with the results of the

dimensional analyses, the mean of the group with NVLD was not significantly different from the means of groups with PDD, MD, or ADHD Combined Type.

Discussion

This study used four existing samples (total N = 8,004) to validate the Colorado Learning Difficulties Questionnaire (CLDQ), a parent-report rating scale designed to screen for learning difficulties in children and adolescents. To the best of our knowledge, the CLDQ is the first parent rating scale designed to assess multiple dimensions of learning difficulties in children and adolescents. Exploratory factor analyses of the CLDQ identified five factors in all four samples, and confirmatory factor analyses indicated that the factor loadings could be equated across samples. In this section we first examine the convergent and discriminant evidence for five CLDQ scales based on the observed factors, then discuss key limitations of the study and areas in which additional studies are needed.

Convergent and discriminant evidence for CLDQ scores

CLDQ Reading scale—Evidence of validity based on internal structure and relations with key external variables is strongest for the CLDQ Reading scale. Factor analyses in all four samples indicated that the six reading-related items loaded strongly on a single factor and did not cross-load on any other factor, and a composite score based on these six items had adequate inter-rater and test-retest reliability. Convergent evidence for the CLDQ Reading scale is provided by significant correlations with standardized measures of reading achievement (overall r = .64). In addition, individuals who met diagnostic criteria for RD scored significantly higher on the Reading scale than on any other CLDQ scale, and the mean of the group with RD was significantly higher than the means of groups with any other disorder. These results provide strong convergent and discriminant evidence for the CLDQ Reading scale.

Results from two large population-based twin studies illustrate the potential utility of the CLDQ Reading scale for research purposes (Hay, Martin, Piek, Levy, & Sheikhi, 2005; Paloyelis, Rijsdijk, Wood, Asherson, & Kuntsi, in press). Because practical constraints precluded the use of individually-administered measures of reading achievement, parent ratings on the CLDQ Reading scale were obtained as part of a larger battery of questionnaires. Results from both studies provided additional support for the internal structure of the CLDQ Reading scale, and behavioral genetic analyses in each sample indicated that the etiology of individual differences in reading was similar to the results obtained by previous twin studies that administered standardized measures of reading achievement (e.g., Bates et al., 2007; Byrne et al., 2007; Petrill et al., 2007). These results suggest that the CLDQ Reading scale may provide a useful research tool to screen for reading difficulties when it is not feasible to administer standardized reading achievement tests.

CLDQ Math scale—Factor analyses in all four samples yielded a math factor, and the CLDQ Math scale was more strongly associated with MD and standardized measures of math achievement than any other CLDQ scale. However, estimates of internal consistency and reliability were lower for the CLDQ Math scale than the Reading scale. These weaker psychometric characteristics may be at least partially explained by the small number of items on the Math scale. In addition, the items on the current Math scale are relatively general, and do not directly assess specific aspects of math performance such as word problems or knowledge of math facts. To address both of these caveats we are currently testing the utility of additional math items in several of the samples. Initial results from the first 70 cases with the new items in the Boulder clinic sample suggest that the addition of

two specific items (*difficulty learning early math facts* and *difficulty with math word problems*) may significantly improve the reliability and predictive power of the current CLDQ Math scale, although a larger sample will be required to fully evaluate the expanded scale. Overall, these results support the validity of scores on the current CLDQ Math scale, but suggest that these additional items may further strengthen the scale and increase its utility for clinical and research purposes.

CLDQ Social Cognition and Social Anxiety scales—Based on previous studies of PDD and NVLD (Hartman, Luteijn, Serra, & Minderaa, 2006; Petti, Voelker, Shore, & Hayman-Abello, 2003; Rourke, 1989), we anticipated that social difficulties would be an important component of the profile of weaknesses exhibited by some children with learning difficulties. The current results support this overall hypothesis, but several findings suggest that it may be useful to examine more specific components of social dysfunction. Factor analyses in all four samples identified a factor characterized by anxiety induced by interpersonal interactions, along with a second factor that included items that reflected weak social awareness or inadequate understanding of social expectations.

Analyses of the external measures provided additional support for the distinction between the CLDQ Social Cognition and Social Anxiety scales. The Social Cognition scale was more strongly related to PDD symptoms, externalizing behavior, rejection by peers, and poor social skills than the Social Anxiety scale, whereas the Social Anxiety scale was more strongly associated with social isolation and symptoms of anxiety disorders. Hartman et al. (2006) reported similar results in a study of the Children's Social Behavior Questionnaire (CSBQ; Luteijn, Luteijn, Jackson, Volkmar, & Minderaa, 2000; Luteijn, Jackson, Volkmar, & Minderaa, 1998), a measure designed to assess dimensions of social behavior that are associated with PDD. In their study, a group with PDD scored significantly higher on the CSBQ Social Understanding subscale than a group with an internalizing disorder and a control group without a diagnosis, whereas the group with an internalizing disorder did not differ significantly from the control group on the Social Understanding scale.

The practical utility of the current CLDQ Social Cognition and Social Anxiety scales is likely to be constrained by psychometric weaknesses. Both scales had lower reliability than the other CLDQ scales, and the final Social Anxiety scale included only three items, one of which cross-loaded with the social cognition items in two of the four factor analyses. Nonetheless, these results suggest that additional research is needed to identify the specific dimensions of social functioning that are impaired in children with LDs or other related developmental difficulties. We are currently testing if the inclusion of additional putative social anxiety and social cognition items further improves the reliability and discriminant evidence for these scales.

CLDQ Spatial scale—Scores on the Spatial scale were significantly associated with external measures of spatial functioning, and were significantly elevated in individuals with NVLD. However, correlations of similar magnitude were also observed between the CLDQ Spatial scale and measures of math and ADHD symptoms, and the mean of the group with NVLD was not significantly different from the mean of groups with ADHD, PDD, or MD. Therefore, the CLDQ Spatial scale appears to be a useful indicator of the spatial difficulties exhibited by individuals with NVLD, ADHD, and other developmental disorders (Forrest, 2004), but it has weaker discriminant evidence than the other scales on the CLDQ.

Clinical utility of the CLDQ scales

To assess the utility of the CLDQ as a screening measure for clinical purposes we are continuing to collect the CLDQ as part of clinical assessments and several ongoing research

studies. As the samples with each specific disorder become sufficiently large, we will be able to test the concordance between categorical cutoff scores on the CLDQ scales and clinical diagnoses of RD, MD, NVLD, and PDD. Preliminary analyses of the current clinic samples suggest that cutoff scores on the CLDQ Reading and Math scales may have sufficient positive and negative predictive power for RD and MD to be clinically useful, and the Social Cognition scale may help to identify individuals with a potential weakness in social functioning that should be assessed in more detail during the assessment.

Although these preliminary results are encouraging, it is important to emphasize that no matter what the final outcome of these future analyses, it will never be appropriate for clinicians to use the CLDQ in isolation to make categorical diagnostic or treatment decisions regarding a specific individual. Instead, by providing an efficient tool to screen for learning difficulties at the beginning of an evaluation, the CLDQ may inform clinical decisions regarding the focus of the assessment, and provide useful supplementary information for case formulation.

Limitations and future directions

A primary strength of the current study is the use of four large samples ascertained in different ways for different purposes. Each sample included a large battery of measures that were used to evaluate the convergent and discriminant evidence for scores on each CLDQ scale. The sample size for most analyses was sufficiently large to provide high power to detect associations between CLDQ scales and key external measures, and also to test whether the magnitude of these associations differed among the CLDQ scales. Findings were generally robust despite potentially important differences between samples in ascertainment, socioeconomic status, ethnicity, age, and the specific battery of external measures completed by the participants. Despite these strengths, this study design also has several inherent weaknesses that should be considered carefully when interpreting the current results and their implications for future research clinical use.

Samples of convenience with missing measures of some constructs—One of the most important limitations of the current study is the fact that none of these samples were recruited for the purpose of evaluating the CLDQ. Because most individuals in the clinic samples were referred for an assessment of ADHD, RD, or other specific learning difficulties, nearly all participants in all four samples completed a standard battery that included measures of intelligence, academic achievement, internalizing and externalizing psychopathology, and social functioning. In contrast, measures of spatial functioning were systematically omitted for some cases in the clinic samples if the referral question and results of other testing did not suggest that spatial difficulties were a specific area of concern.

Two sets of secondary analyses were conducted to test whether the omission of spatial measures from this subset of cases biased analyses of the associations between these measures and the CLDQ Spatial scale. The first set of analyses directly compared the subset of the clinic samples that completed the spatial measures (N = 589) to the group that did not complete these tasks (N = 482). The CLDQ Spatial score of the group was significantly higher in the group that completed the spatial measures, but the effect size was small (d = .19), and the two groups did not differ on the other four CLDQ scales or any other external measures. The second set of analyses compared results in the four samples to test if a different pattern emerged in the clinic and community samples. Correlations between the CLDQ Spatial scale and the external measures of spatial functioning were nearly identical in all samples (r = .27 – .32).

Taken together, these results suggest that the omission of the spatial measures from a subset of the cases in the clinic samples had minimal impact on the overall pattern of results. Nonetheless, future studies of clinic samples could provide a useful extension of the current research by administering a standard test battery to all participants that includes multiple measures of each of the constructs assessed by the CLDQ.

Skewed scores and restricted range on the CLDQ scales or external variables—A second concern related to the use of samples of convenience is the possibility that the distribution of some measures might violate statistical assumptions of normality. For example, low scores on a CLDQ scale could be underrepresented in a clinic sample if most cases seen by the clinic have difficulties related to a specific scale (e.g., CLDQ Reading scores in a clinic sample with a high proportion of RD cases). However, skewness and kurtosis were within normal limits (i.e., absolute value less than 1) for all CLDQ scales and external measures in the clinic samples, suggesting that correlations were not attenuated by a restricted range of scores. Distributions of CLDQ scores in the community samples were characterized by mild positive skew due to the large number of individuals with no learning difficulties (skewness = 1.1 – 1.6), but skewness was adequately reduced after the data were suitably transformed. Most importantly, the pattern of results was extremely similar in the four samples for all primary analyses, suggesting that any violations of statistical assumptions did not have a major impact on the results.

Use of the questionnaire for case formulation in the clinic samples—Although the final clinical diagnosis was based primarily on other information obtained during the assessment, parent ratings on the CLDQ were one component of the clinical data used for case formulation in the Denver clinic sample (in the Boulder clinic the CLDQ was included solely for research purposes to avoid this potential confound). If high ratings on the CLDQ strongly influenced the final diagnosis that a child received in the Denver sample, the mean CLDQ score of groups with specific diagnoses could be biased upward. Consistent with this hypothesis, the effect size for the RD group on the CLDQ Reading scale was higher in the Denver Clinic sample (d_w = 1.92) than the community samples (d_w = 1.64). However, the CLDQ Reading score in the Boulder Clinic sample (d_w = 2.08) was even higher than the score in the Denver clinic. Further, even in the community samples the effect size for the RD group was substantially larger on the CLDQ Reading scale than any other CLDQ scale, and there were no other significant differences between the clinical and community samples for any other comparison. Overall, this pattern of results suggests that although the means of the RD group on the CLDQ Reading scale were significantly higher in the clinic samples, this difference was not a specific consequence of the use of CLDQ scores as part of the overall case formulation.

Small item pool for some constructs—The initial item pool for the CLDQ was developed to screen for a range of common parental concerns as part of a lengthy developmental history questionnaire completed by parents at the beginning of their child's assessment. Therefore, it was not feasible to ask parents to complete the large number of items (i.e., 200 – 300) that are often included in an initial pool of items when the primary goal of a study is the development and validation of a new measure (e.g., Achenbach & Rescorla, 2001; Lahey et al., 2004; Reynolds & Kamphaus, 2004). Due to the relatively small size of the initial item pool (46 items) and the exclusion of over half of the initial items for theoretical and psychometric reasons, the CLDQ Math and Social Anxiety Scales included only three items. As noted previously, additional items are currently being evaluated to evaluate whether their inclusion improves the internal structure and convergent and discriminant evidence for these scales.

Relevant constructs not measured by the CLDQ—The CLDQ does not assess several domains that are often correlated with learning difficulties, including written and spoken language, motor skills, and processing speed (e.g., Bishop & Snowling, 2004; Pitcher, Piek, & Barrett, 2002; Shanahan et al., 2006). In addition, although several items on the initial scale were designed to measure memory difficulties, these items were dropped from the final scale due to weak psychometric characteristics or absence of loadings above .40 on any factor in the EFA.

Small samples with some clinical disorders—The two community samples were recruited for studies of RD, ADHD, or both disorders. The assessment clinics received a more diverse range of referral questions, but a majority of the evaluations also focused on questions regarding RD, ADHD, and related disorders. Therefore, in comparison to the samples with RD or ADHD, a smaller number of participants met criteria for less common disorders such as MD, NVLD, and PDD. Moreover, sample sizes were too small to examine potentially important distinctions between disorders within these broad diagnostic clusters, such as Autistic Disorder versus Asperger's Disorder. Future studies of the relation between CLDQ scales and larger samples of individuals with these disorders would provide a useful extension of the present results.

Conclusions

Exploratory and confirmatory factor analyses of the Colorado Learning Difficulties Questionnaire (CLDQ) revealed five correlated but separable dimensions of learning difficulties in children and adolescents. Results provide strong convergent and discriminant evidence for scores on a 6-item Reading scale, and suggest that this scale may provide a useful screening measure for reading difficulties in both research and clinical settings. Results are also promising for scales that assess math, social anxiety, social cognition, and spatial difficulties, but additional research is needed to address specific weaknesses identified in each of these scales.

Acknowledgments

This research was supported by grants from the National Institute of Child Health and Human Development (P50 HD27802) and the National Institute of Mental Health (R01 MH 62120, R01 MH 63941, and R01 MH 70037), and by annual Outreach grants from the University of Colorado, Boulder from 2004 – 2010. The authors were also supported by NIH grants R01 HD 47264, R01 DC 05190, R01 HD38526 during the preparation of this report.

References

Achenbach, TM.; Rescorla, LA. Manual for ASEBA School-age Forms and Profiles. University of Vermont Research Center for Children, Youth, and Families; Burlington, VT: 2001.

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision. American Psychiatric Association; Washington, DC: 2000.

Antshel KM, Khan FM. Is There an Increased Familial Prevalence of Psychopathology in Children With Nonverbal Learning Disorders? Journal of Learning Disabilities. 2008; 41:208–217. [PubMed: 18434288]

Barkley, RA.; Murphy, K. Attention-deficit hyperactivity disorder: A clinical workbook. 2nd ed.. Guilford Press; New York, NY: 1998.

Bates TC, Castles A, Luciano M, Wright MJ, Coltheart M, Martin NG. Genetic and environmental bases of reading and spelling: A unified genetic dual route model. Reading and Writing. 2007; 20:147–171.

Beery, KE.; Buktenica, NA. Beery-Buktenica Developmental Test of Visual-Motor Integration, Fourth Edition. Modern Curriculum Press; Parsippany,NJ: 1997.

Bentler PM. Comparative Fit Indexes in Structural Models. Psychological Bulletin. 1990; 107:238–246. [PubMed: 2320703]

Bishop DVM, Snowling MJ. Developmental dyslexia and specific language impairment: Same or different? Psychological Bulletin. 2004; 130:858–886. [PubMed: 15535741]

Boetsch EA, Green PA, Pennington BF. Psychosocial correlates of dyslexia across the life span. Development and Psychopathology. 1996; 8:539–562.

Browne, MW.; Cudeck, R. Alternative ways of assessing model fit. In: Bollen, KA.; Long, JS., editors. Testing structural equation models. Sage; Newbury Park, CA: 1993. p. 136-162.

Byrne B, Samuelsson S, Wadsworth SJ, Hulslander J, Corley R, DeFries JC, et al. Longitudinal twin study of early literacy development: preschool through grade 1. Reading and Writing. 2007; 20:77–102.

Chen F, Curran PJ, Bollen KA, Kirby J, Paxton P. An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. Sociological Methods and Research. 2008; 36:462–494. [PubMed: 19756246]

Conners CK, Sitarenios G, Parker JD, Epstein JN. Revision and restandardization of the Conners Teacher Rating Scale (CTRS-R): factor structure, reliability, and criterion validity. Journal of Abnormal Child Psychology. 1998; 26:279–291. [PubMed: 9700520]

Daniel SS, Walsh AK, Goldston DB, Arnold EM, Reboussin BA, Wood FB. Suicidality, school dropout, and reading problems among adolescents. Journal of Learning Disabilities. 2006; 39:507–514. [PubMed: 17165618]

DeFries JC, Filipek PA, Fulker DW, Olson RK, Pennington BF, Smith SD, et al. Colorado Learning Disabilities Research Center. Learning Disabilities: A Multidisciplinary Journal. 1997; 8:7–19.

DerSimonian R, Laird N. Metaanalysis in Clinical-Trials. Controlled Clinical Trials. 1986; 7:177–188. [PubMed: 3802833]

Dishion, T. The peer context of troublesome child and adolescent behavior. In: Leone, PE., editor. Understanding troubled and troubling youth. Sage; Newbury Park, CA: 1990.

Dunn, LM.; Markwardt, FC. Examiner's Manual: Peabody Individual Achievement Test. American Guidance Service; Circle Pines, MN: 1970.

DuPaul, GJ.; Power, TP.; Anastopoulos, AD.; Reid, R. ADHD Rating Scale - IV. Guilford Press; New York: 1998.

Fan XB, Thompson B, Wang L. Effects of sample size, estimation method, and model specification on structural equation modeling fit indexes. Structural Equation Modeling. 1999; 6:56–83.

Forrest BJ. The utility of math difficulties, internalized psychopathology, and visual-spatial deficits to identify children with the nonverbal learning disability syndrome: Evidence for a visualspatial disability. Child Neuropsychology. 2004; 10:129–146. [PubMed: 15590491]

Gadow, KD.; Sprafkin, J. Childhood Symptom Inventory 4: Norms Manual. Checkmate Plus; Stony Brook, NY: 1997a.

Gadow, KD.; Sprafkin, J. Early Childhood Inventory 4: Norms Manual. Checkmate Plus; Stony Brook, NY: 1997b.

Gadow, KD.; Sprafkin, J. Adolescent Symptom Inventory 4: Norms Manual. Checkmate Plus; Stony Brook, NY: 1998.

Goldston DB, Walsh A, Mayfield AE, Reboussin B, Sergent DS, Erkanli A, et al. Reading problems, psychiatric disorders, and functional impairment from mid- to late adolescence. Journal of the American Academy of Child and Adolescent Psychiatry. 2007; 46:25–32. [PubMed: 17195726]

Gross-Tsur V, Manor O, Shalev RS. Developmental dyscalculia: prevalence and demographic features. Developmental Medicine and Child Neurology. 1996; 38:25–33. [PubMed: 8606013]

Hartman CA, Luteijn E, Serra M, Minderaa R. Refinement of the children's social behavior questionnaire (CSBQ): An instrument that describes the diverse problems seen in milder forms of PDD. Journal of Autism and Developmental Disorders. 2006; 36:325–342. [PubMed: 16617405]

Hay D, Martin N, Piek J, Levy F, Sheikhi A. A parent-report approach to the genetics of the relationship between ADHD and reading problems. Behavior Genetics. 2005; 35:805.

Hedges, LV.; Olkin, I. Statistical methods for meta-analysis. Academic Press; Orlando, FL: 1985.

Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Statistics in Medicine. 2002; 21:1539–1558. [PubMed: 12111919]

Jastak, S.; Wilkinson, GS. Wide Range Achievement Test, Revised: Administration Manual. Wilmington, DE: 1984.

Klin A, Volkmar FR, Sparrow SS, Cicchetti DV, Rourke BP. Validity and Neuropsychological Characterization of Asperger Syndrome - Convergence with Nonverbal Learning-Disabilities Syndrome. Journal of Child Psychology and Psychiatry and Allied Disciplines. 1995; 36:1127–1140.

Kovacs, M. The Children's Depression Inventory. Multi-health Systems, Inc.; New York: 1988.

Lahey BB, Applegate B, Waldman ID, Loft JD, Hankin BL, Rick J. The structure of child and adolescent psychopathology: generating new hypotheses. Journal of Abnormal Psychology. 2004; 113:358–385. [PubMed: 15311983]

Luteijn E, Luteijn F, Jackson S, Volkmar F, Minderaa R. The children's Social Behavior Questionnaire for milder variants of PDD problems: evaluation of the psychometric characteristics. Journal of Autism and Developmental Disorders. 2000; 30:317–330. [PubMed: 11039858]

Luteijn EEF, Jackson SAE, Volkmar FR, Minderaa RB. Brief report: The development of the Children's Social Behavior Questionnaire: Preliminary data. Journal of Autism and Developmental Disorders. 1998; 28:559–565. [PubMed: 9932242]

Maughan B, Rowe R, Loeber R, Stouthamer-Loeber M. Reading problems and depressed mood. Journal of Abnormal Child Psychology. 2003; 31:219–229. [PubMed: 12735404]

McGee R, Prior M, Willams S, Smart D, Sanson A. The long-term significance of teacher-rated hyperactivity and reading ability in childhood: findings from two longitudinal studies. Journal of Child Psychology and Psychiatry. 2002; 43:1004–1017. [PubMed: 12455922]

Meyers, JE.; Meyers, KR. Rey Complex Figure Test and Recognition Trial. Psychological Assessment Resources; Odessa, FL: 1995.

Muthen, LK.; Muthen, BO. Mplus User's Guide: Fourth Edition. Muthen and Muthen; Los Angeles, CA: 2009.

Paloyelis Y, Rijsdijk F, Wood AC, Asherson P, Kuntsi J. The genetic association between ADHD symptoms and reading difficulties: The role of inattentiveness and IQ. Journal of Abnormal Child Psychology. in press.

Petrill SA, Deater-Deckard K, Thompson LA, Schatschneider C, Dethorne LS, Vandenbergh DJ. Longitudinal genetic analysis of early reading: The Western Reserve Reading Project. Reading and Writing. 2007; 20:127–146. [PubMed: 19829751]

Petti VL, Voelker SL, Shore DL, Hayman-Abello SE. Perception of nonverbal emotion cues by children with nonverbal learning disabilities. Journal of Developmental and Physical Disabilities. 2003; 15:23–36.

Pitcher TM, Piek JP, Barrett NC. Timing and force control in boys with attention deficit hyperactivity disorder: subtype differences and the effect of comorbid developmental coordination disorder. Human Movement Science. 2002; 21:919–945. [PubMed: 12620726]

Reich, W.; Welner, Z.; Herjanic, B. Diagnostic Interview for Children and Adolescents - IV. Multi-Health System, Inc.; North Towanda Falls, NY: 1997.

Rey A. L'examen psychologique dans les cas d'encephalopathie tramatique. Archives of Psychology. 1941; 28:286–340.

Reynolds, CR.; Kamphaus, RW. Behavior Assessment System for Children. American Guidance Service; Circle Pines, MN: 1992.

Reynolds, CR.; Kamphaus, RW. Behavior Assessment System for Children, Second Edition. American Guidance Service; Circle Pines, MN: 2004.

Rourke, BP. Nonverbal Learning Disabilities: The Syndrome and the Model. Guilford; New York: 1989.

Rutter M, Caspi A, Fergusson D, Horwood LJ, Goodman R, Maughan B, et al. Sex differences in developmental reading disability - New findings from 4 epidemiological studies. Journal of the American Medical Association. 2004; 291:2007–2012. [PubMed: 15113820]

Schnoes C, Reid R, Wagner M, Marder C. ADHD among students receiving special education services: A national survey. Exceptional Children. 2006; 72:483–496.

Schumacker, RE.; Lomax, RG. A beginner's guide to structural equation modeling, second edition. Lawrence Erlbaum Associates; Mahwah,NJ: 2004.

Semrud-Clikeman M, Biederman J, Sprich-Buckminster S, Lehman BK, Faraone SV, Norman D. Comorbidity between ADDH and learning disability: a review and report in a clinically referred sample. Journal of the American Academy of Child and Adolescent Psychiatry. 1992; 31:439–448. [PubMed: 1592775]

Shanahan MA, Pennington BF, Yerys BE, Scott A, Boada R, Willcutt EG, et al. Processing speed deficits in attention deficit/hyperactivity disorder and reading disability. Journal of Abnormal Child Psychology. 2006; 34:585–602. [PubMed: 16850284]

Shaywitz SE, Shaywitz BA, Fletcher JM, Escobar MD. Prevalence of reading disability in boys and girls. Results of the Connecticut Longitudinal Study. Journal of the American Medical Association. 1990; 264:998–1002. [PubMed: 2376893]

Thurstone, TG. Examiners' manual: Primary Mental Abilities. Science Research Associates; Chicago, IL: 1963.

Trzesniewski KH, Moffitt TE, Caspi A, Taylor A, Maughan B. Revisiting the association between reading achievement and antisocial behavior: new evidence of an environmental explanation from a twin study. Child Development. 2006; 77:72–88. [PubMed: 16460526]

Wechsler, D. Manual for the Wechsler Intelligence Scale for Children, Revised. The Psychological Corporation; New York, NY: 1974.

Wechsler, D. Manual for the Wechsler Intelligence Scale for Children, Third Edition. The Psychological Corporation; San Antonio,TX: 1991.

Wechsler, D. Manual for the Wechsler Intelligence Scale for Children, Fourth Edition. The Psychological Corporation; San Antonio,TX: 2003.

Wiederhold, JL.; Bryant, BR. Gray Oral Reading Test - Third Edition. Pro-Ed; Austin, TX: 1993.

Wiederhold, JL.; Bryant, BR. Gray Oral Reading Test - Fourth Edition. Pro-Ed; Austin, TX: 2001.

Wilkinson, GS. Examiner's Manual: Wide Range Achievement Test - 3. Jastak Associates - Wide Range, Inc.; Wilmington, DE: 1993.

Willcutt EG, Betjemann RS, Pennington BF, Olson RK, DeFries JC, Wadsworth SJ. Longitudinal study of reading disability and attention-deficit/hyperactivity disorder: Implications for education. Mind, Brain, and Education. 2007; 4:181–192.

Willcutt EG, Pennington BF. Comorbidity of reading disability and attention-deficit/hyperactivity disorder: differences by gender and subtype. Journal of Learning Disabilities. 2000a; 33:179–191. [PubMed: 15505947]

Willcutt EG, Pennington BF. Psychiatric comorbidity in children and adolescents with reading disability. Journal of Child Psychology and Psychiatry. 2000b; 41:1039–1048. [PubMed: 11099120]

Willcutt EG, Pennington BF, Olson RK, Chhabildas N, Hulslander J. Neuropsychological analyses of comorbidity between reading disability and attention deficit hyperactivity disorder: in search of the common deficit. Developmental Neuropsychology. 2005; 27:35–78. [PubMed: 15737942]

Woodcock, RW.; Johnson, MB. Manual for the Woodcock-Johnson Psycho-educational Battery - Revised (WJ-R). Riverside Publishing Company; Chicago, IL: 1990.

Woodcock, RW.; McGrew, KS.; Mather, N. Woodcock-Johnson III Tests of Achievement. Riverside Publishing; Itasca,IL: 2001.

Table 1

Description of samples

	Denver Clinic	Boulder Clinic	Twin Sample	Community Sample
		Descriptive Characteristics		
Total N	954	179	1,840	5,031[a]
Type of sample	Clinic	Clinic	Community	Community
Selection of sample	Consecutive Cases	Consecutive Cases	Enriched for RD & ADHD	Unselected[a]
Percent female[b]	32.2%[a]	33.5%[a]	50.8%[b]	51.6%[b]
Age range	6 – 18	6 – 18	8 – 18	6 – 14
Mean Age (SD)[b]	11.0 (3.5)	10.9 (3.2)	11.2 (2.6)	10.8 (2.9)
% White, non-hispanic[b]	89.0%[a]	85.9%[a]	83.3%[a]	58.7%[b]
Father years of education[b]	16.9 (2.9)[a]	17.3 (2.9)[a]	14.7 (2.7)[b]	13.4 (2.7)[c]
Mother years of education[b]	16.2 (2.5)[a]	16.7 (3.4)[a]	14.4 (2.4)[b]	13.9 (2.9)[c]

	Diagnoses							
	N	**%**	**N**	**%**	**N**	**%**	**N**	**%**
No Diagnosis	44	4.6%	9	5.0%	868	47.2%	4,547	90.4%
Reading Disorder	422	44.2%	69	38.5%	496	27.0%	54[d]	5.3%
Math or Nonverbal LD	36	3.8%	19	10.6%	131[d]	7.1%	41[d]	4.0%
ADHD Inattentive Type	135	14.2%	50	27.9%	273	14.8%	246	4.9%
ADHD Combined Type	277	29.0%	67	37.4%	133	7.2%	238	4.7%
Disruptive Disorders	21	2.2%	15	8.4%	258	14.0%	147[f]	20.5%
Mood Disorder	94	9.9%	14	7.8%	82	4.5%	31[f]	4.3%
Anxiety Disorder	22	2.3%	11	6.1%	309	16.8%	22[f]	3.1%
Pervasive Develop. Disorder	15	1.6%	10	5.6%	--[g]	--[g]	24	0.5%
Other diagnosis	127[h]	13.3%	9[i]	5.0%	--	--	--	--

Note. --indicates variable not assessed.

[a] An initial unselected sample of 5,031 children was screened for ADHD, and a random sample of individuals with and without ADHD were recruited for the detailed assessment.

[b] Different subscripts indicate a significant difference between samples (P < .01)

[c] Ns indicate total number of participants with each diagnosis. Because many participants met criteria for more than one disorder, the sum of the Ns for all diagnoses is more than the total number of participants.

[d] Comorbid LDs were assessed in 1,048 children who completed achievement testing as part of the detailed assessment.

[d] Math achievement assessed in 1,014 individuals.

[f] Disorder assessed in the 716 participants whose parents completed a diagnostic interview.

[g] Exclusion criterion.

[h] Tourettes / Tic Disorder (N = 14), Cognitive Disorder, not otherwise specified (N = 54), Obsessive-compulsive disorder (N = 4), Phonological / Speech Disorders (N = 17), Receptive / Expressive Language Disorder (N = 38).

[i] Tourettes / Tic Disorder (N = 1), Obsessive-compulsive disorder (N = 1), Receptive / Expressive Language Disorder (N = 5).

Table 2

The 46 items on the initial CLDQ

Items included on the final scale [a]

1. Difficulty with spelling
2. Difficulty learning letter names
3. Difficulty learning phonics (sounding out words)
4. Read slowly
5. Read below grade or expectancy level
6. Required extra help in school because of problems in reading and spelling
7. Poor understanding of interpersonal space
8. Difficulty knowing how others are reacting
9. Has trouble understanding how others are feeling
10. Makes comments that show a lack of understanding of social situations, such as inappropriate jokes or insensitive remarks
11. Difficulty making or keeping friends
12. Isolates self in social situations
13. Feels anxious or out-of-place in new social situations
14. Handwriting is spatially disorganized
15. Papers look disorganized or messy
16. On arithmetic problems, has difficulty keeping the numbers lined up in columns
17. Drawings look immature for her/his age
18. Worse at math than at reading and spelling
19. Makes careless errors in math, such as adding when the sign indicates subtraction
20. Trouble learning new math concepts such as carrying or borrowing

Items dropped because they assess behaviors related to ADHD

21. Leaves things unfinished, like starting a game and then running off to do something else
22. Gets into trouble because he/she rushes into doing things without thinking about what could happen
23. Rush through assignments without checking them
24. Teachers often have to tell her/him what to do after the rest of the class has started
25. When playing games or lining up for class, tries to get in before his/her turn or pushes ahead in line
26. People keep telling your child to sit still
27. People keep telling your child to stop fidgeting
28. People tell you that your child is (was) always on the go, as if he/she was driven by a motor
29. Difficulty keeping his/her mind on things he/she enjoys, such as reading a story or watching TV
30. Trouble sticking to what he/she is told to do if there are noises or people moving around in the room
31. Loses or misplaces things more than others

Items dropped because they assess behaviors related to depression

32. Sad or unhappy
33. Seems to have less energy than others his/her age
34. Seems to feel at fault when something goes wrong
35. Says things like “I wish I were dead” or “I wish I had never been born”
36. Expressed more definite suicidal thoughts or wishes

Items dropped because they did not load above .40 on any factor or had weak reliability.

37. Difficulty remembering things that others her/his age seem to remember
38. Seems to forget the content of TV shows or movies
39. Difficulty remembering things that happened to her/him in the past
40. Difficulty with new skills involving small muscle coordination, such as cutting or writing
41. Difficulty remembering names, lists, phone numbers, or complex instructions
42. Inappropriate eye contact when interacting with others
43. Slow in developing basic self-help skills, such as independent dressing or use of eating utensils
44. More difficulty with puzzles than other children
45. When learning his/her way around new places, seems to lack a good sense of direction
46. Difficulty learning the days of the week or months of the year

[a] Items included the initial stem “Does / did your child have…”.

Table 3

Principal axis factor analyses and reliability of the final pool of CLDQ items

	Reliability		Primary factor loading in the four samples[a]
	Inter-rater[b]	Test-retest[c]	Mean (low, high)
Factor 1: Reading			
1. Difficulty with spelling	.67	.69	.70 (.60, .77)
2. Difficulty learning letters	.58	.60	.65 (.55, .75)
3. Difficulty learning phonics	.67	.68	.80 (.75, .87)
4. Reads slowly	.74	.69	.86 (.80, .91)
5. Reads below grade level	.75	.75	.87 (.83, .90)
6. Required extra reading help	.78	.73	.85 (.80, .89)
Factor 2: Social cognition			
7. Poor understanding of interpersonal space	.47	.56	.72 (.70, .73)
8. Difficulty knowing how others are reacting	.45	.57	.87 (.85, .91)
9. Difficulty understanding the feelings of others	.44	.60	.83 (.79, .88)
10. Comments lack social understanding	.46	.61	.74 (.69, .78)
Factor 3: Social anxiety			
11. Difficulty making or keeping friends	.57	.64	.66[d] (.56, .72)
12. Isolates self in social situations	.53	.52	.90 (.82, .97)
13. Feels anxious in new social situations	.46	.53	.73 (.70, .74)
Factor 4: Spatial			
14. Handwriting is spatially disorganized	.54	.66	.83 (.70, .89)
15. Papers look disorganized or messy	.62	.66	.81 (.74, .90)
16. Trouble keeping numbers in columns	.54	.59	.74 (.70, .76)
17. Drawings look immature for her/his age	.56	.64	.61 (.56, .68)
Factor 5: Math			
18. Worse at math than at reading and spelling	.54	.61	.86 (.81, .89)
19. Makes careless errors in math	.49	.59	.74 (.66, .87)
20. Trouble learning new math concepts	.60	.70	.79 (.72, .87)

[a] Pattern matrix loadings from principal axis factor analyses with oblimin rotation. Loadings are mean of the four samples weighted by sample size. The full pattern and structure matrix for all samples are available from the lead author upon request.

[b] Correlation between mother and father ratings in the twin sample (N = 1,124).

[c] One-year test-retest reliability of maternal ratings in a subset of the community screening sample (N = 524). All inter-rater and test-retest correlations are significant (P < .001).

[d] item cross-loaded in the Denver clinic sample (.32) and the community sample (.43).

Table 4

Reliability of CLDQ scores and correlations between scales

		Reliability of composite scores			Correlations between composite scores[a]			
					Reading	Social Cognition	Social Anxiety	Spatial
Composite Score	**Estimated population mean and SD[b]**	**Cronbach's α Mean (low, high)**	**Inter-rater reliability[c]**	**Test-retest reliability[d]**	**r_w [95% CI]**	**r_w [95% CI]**	**r_w [95% CI]**	**r_w [95% CI]**
Reading	1.79 (0.94)	.90 (.88 – .93)	.83	.81	--			
Social Cognition	1.54 (0.73)	.86 (.83 – .88)	.53	.71	.28 [.15, .40]	--		
Social Anxiety	1.60 (0.75)	.82 (.79 – .83)	.59	.68	.22 [.09, .33]	.56 [.52, .59]	--	
Spatial	1.77 (0.92)	.85 (.82 – .88)	.69	.76	.37 [.25, .48]	.38 [.32, .44]	.33 [.28, .37]	--
Math	1.73 (0.88)	.80 (.76 – .82)	.63	.73	.32 [.22, .42]	.25 [.18, .32]	.23 [.18, .28]	.47 [.41, .53]

[a] r_W = overall correlation and 95% confidence interval were estimated with random effects models described by DerSimonian & Laird (1986). Total N = 7,634.

[b] The population mean and SD was estimated from the unselected sample of participants in the community study.

[c] Correlation between mother and father ratings in the twin sample (N = 1,124).

[d] One-year test-retest reliability of maternal ratings in a subset of the community sample (N = 524).

Table 5

Measures used to evaluate the external validity of the CLDQ scales

	Denver clinic (Total N = 954)		Boulder clinic (Total N = 179)		Twins (Total N = 1,840)		Community Sample (Total N = 5,031)[a]	
Construct	**Measure**	**N**	**Measure**	**N**	**Measure**	**N**	**Measure**	**N**
Reading								
Single-word reading	WJ Letter Word ID	928	WJ-III Letter-Word ID	177	PIAT Read Rec	1,840	WJ-III Letter Word ID	1,564[a]
Reading Comprehension	GORT-III	847	GORT-IV	161	PIAT Reading Comp	1,840	--	--
Math								
Calculations	WJ Calculations	928	WJ-III Calculations	177	WRAT Math	1,660	WJ-III Calculations	1,564[a]
Word Problems	WJ Applied Prob.	867	WJ-III Applied Prob.	177	PIAT Math	1,656	--	--
Social functioning								
Social Problems	ASEBA	642[b]	BASC-II	178	ASEBA	1,815	BASC	2,224[c]
Withdrawn Behaviors	ASEBA	642[b]	BASC-II	178	ASEBA	1,815	BASC	2,224[c]
Social Skills	--	--	BASC-II	178	--	--	BASC	2,224[c]
Sociometric ratings								
Liked by peers	--	--	Number of friends	172	Dishion (1990)	1,474[d]	Dishion (1990)	5,025
Disliked by peers	--	--	--	--	Dishion (1990)	1,474[d]	Dishion (1990)	5,025
Ignored by peers	--	--	--	--	Dishion (1990)	1,474[d]	Dishion (1990)	5,021
Spatial functioning								
Block Design	WISC-III	803[h]	WISC-IV	173[g]	WISC-R	1,840	WISC-III	1,564[a]
Other spatial	ROCFT Copy	324[d]	ROCFT Copy	121[d]	ROCFT Copy	308[h]	Spatial Span	214[b]
	DTVMI	478[d]	--	--	PMA Spatial	1,840	--	--
Psychopathology								
Anxiety	ASEBA[e]	642[b]	BASC-II	174	ASEBA[e]	1,815	BASC	2,224[a]
	--	--	CSI / ASI GAD	174	DICA-IV GAD / SAD	1,840	DICA-IV GAD / SAD	716[f]
Depression	ASEBA[e]	642[b]	BASC-II	174	ASEBA[e]	1,815	BASC	2,224
	--	--	CSI / ASI MDD	175	DICA-IV MDD	1,840	DICA-IV MDD	716[f]
ADHD	DBRS	771[b]	CSI / ASI	175	DBRS	1,815	DBRS	5,031

	Denver clinic (Total N = 954)		Boulder clinic (Total N = 179)		Twins (Total N = 1,840)		Community Sample (Total N = 5,031)[a]	
Construct	**Measure**	**N**	**Measure**	**N**	**Measure**	**N**	**Measure**	**N**
Aggressive Behavior	ASEBA	642[b]	BASC-II	175	ASEBA	1,815	BASC	2,224[c]
Delinquent Behavior	ASEBA	642[b]	BASC-II	175	ASEBA	1,815	BASC	2,224[c]
Pervasive Develop. Dis.	Clinical Diagnosis	15	CSI / ASI[i]	175	Exclusion Criterion	--	Parent Report	24

Note. Measures with no superscripts are part of the standard test battery. Any differences between the total sample size and the sample with the measure indicates missing data.

[a] Measure completed only by the selected sample of 502 participants with ADHD, 532 siblings of the probands, and a comparison group of 530 participants without ADHD.

[b] Added to the standard battery after the initial inception of data collection.

[c] The BASC was only included in the standard screening packet in a subset of school districts to comply with requests of district administrators

[d] Administered to a subset of cases depending on presenting concerns.

[e] anxious/depressed subscale.

[f] The DICA-IV was administered to ADHD and control probands, but not siblings.

[g] A Wechsler IQ test was administered as part of the standard test battery unless the child had recently completed the test as part of another evaluation.

[h] Administered previously as part of the standard test battery, then discontinued.

[i] CSI and ASI ratings of PDD symptoms were used for correlational analyses, and PDD diagnoses were made based on all available clinical information. LWID = Letter Word Identification; ASEBA = parent (Child Behavior Checklist) and teacher (Teacher Report Form) ratings on the Achenbach System of Empirically Based Assessment (Achenbach & Rescorla, 2001); ASI = Adolescent Symptom Inventory (Gadow & Sprafkin, 1998); BASC = parent and teacher ratings on the Behavior Assessment System for Children (BASC: Reynolds & Kamphaus, 1992; BASC-II: Reynolds & Kamphaus, 2004); CSI = Child Symptom Inventory (Gadow & Sprafkin, 1997a); DBRS = Disruptive Behavior Rating Scale (Barkley & Murphy, 1998); DICA-IV = DSM-IV Diagnostic Interview for Children and Adolescents (Reich, Welner, & Herjanic, 1997); GORT = Gray Oral Reading Test (GORT - III = third edition, Wiederhold & Bryant, 1993; GORT - IV = fourth edition, Wiederhold & Bryant, 2001); PIAT = Peabody Individual Achievement Test (Dunn & Markwardt, 1970); PMA Spatial = Primary Mental Abilities - Spatial Relations subtest (Thurstone, 1963), ROCFT = Rey-Osterreith Complex Figure Test (Rey, 1941; scoring procedures are described by Meyers & Meyers, 1995); DTVMI = Developmental Test of Visual-motor Integration (Beery & Buktenica, 1997); WISC-R, WISC-III, and WISC-IV = Wechsler Intelligence Scale for Children (WISC-R = revised, Wechsler, 1974; WISC-III = Third Edition, Wechsler, 1991; WISC-IV = Fourth Edition, Wechsler, 2003); WJ-R and WJ-III = Woodcock-Johnson Tests of Achievement (WJ-R = revised, Woodcock & Johnson, 1990; WJ - III = third edition, Woodcock, McGrew, & Mather, 2001), WRAT = Wide Range Achievement Test (Jastak & Wilkinson, 1984; Wilkinson, 1993).

Psychol Assess. Author manuscript; available in PMC 2012 September 1.

Table 6

Correlations between CLDQ scales and measures used for external validation

Measures of Related Constructs	Samples[b]	Total N	Correlations between CLDQ scales and the external measures[a]				
			Reading r_w [95% CI]	Math r_w [95% CI]	Social Cognition r_w [95% CI]	Social Anxiety r_w [95% CI]	Spatial r_w [95% CI]
Academic achievement							
Reading composite	1 – 4	4,518	.64 [.60, .68]**	.19 [.10, .28]**	.16 [.10, .22]**	.07 [−.03, .16]	.17 [.10, .23]**
Math composite	1 – 4	4,518	.34 [.28, .40]**	.44 [.40, .48]**	.12 [.05, .19]*	.06 [.02, .10]*	.26 [.20, .32]**
Social functioning							
Social rejection composite	1 – 4	7,686	.20 [.16, .24]**	.25 [.22, .28]**	.44 [.35, .53]**	.28 [.26, .31]**	.23 [.19, .27]**
Social isolation composite	1 – 4	7,686	.17 [.12, .21]**	.22 [.17, .27]**	.27 [.24, .30]**	.46 [.36, .56]**	.19 [.13, .25]**
Social strengths composite	2 – 4	6,436	.22 [.20, .25]**	.12 [.08, .16]**	.41 [.32, .49]**	.26 [.20, .31]**	.24 [.21, .26]**
Spatial functioning							
Spatial composite	1 – 4	4,511	.16 [.11, .21]**	.35 [.30, .40]**	.10 [.07, .13]*	.07 [.04, .10]*	.30 [.27, .33]**
Psychopathology							
Anxiety composite	1 – 4	7,686	.13 [.11, .15]**	.14 [.11, .17]**	.23 [.16, .30]**	.36 [.32, .40]**	.12 [.07, .16]**
Depression symptoms	2 – 4	3,744	.20 (.17, .22)**	.19 [.15, .23]**	.28 (.24, .31)**	.32 (.26, .37)**	.22 (.19, .25)**
Externalizing Behaviors	1 – 4	7,686	.11 (.07, .15)**	.09 [.01,. 17]*	.43 (.39, .47)**	.23 (.19, .27)**	.12 (.08, .16)**
Inattention symptoms	1 – 4	7,815	.46 (.40, .51)**	.32 [.22, .42]**	.52 (.41, .63)**	.32 (.27, .38)**	.56 (.50, .62)**
Hyperactivity - Impulsivity symptoms	1 – 4	7,815	.28 (.21, .34)**	.16 [.10, .22]**	.47 (.34, .60)**	.20 (.14, .26)**	.36 (.29, .43)**
Pervasive Dev. Disorder symptoms	2	179	.08 (−.07, .22)	.24 [.11, .37]**	.64 (.54, .74)**	.41 (.27, .54)**	.16 (.01, .30)*

[a] All measures were transformed so that a positive correlation indicates a relation between higher CLDQ ratings and more severe difficulties on the external measure, regardless of the original scaling of the measure. Point estimates are estimated overall correlation based on a random effects model (r_W; DerSimonian and Laird, 1986).

[b] 1. Denver clinic, 2. Boulder clinic, 3. twin sample, 4. community sample. PDD = Pervasive Developmental Disorder.

* = P < .01

** = P < .001.

Table 7

Scores of groups with clinical diagnoses on the five CLDQ scales

			Effect size of the clinical group versus the estimated population mean on each CLDQ scale[a]				
Diagnosis	**Samples[b]**	**Total N**	**Reading d_w [95% CI]**	**Math d_w [95% CI]**	**Social Cognition d_w [95% CI]**	**Social Anxiety d_w [95% CI]**	**Spatial d_w [95% CI]**
Reading Disorder	1 – 4	1,041	1.80 [1.42, 2.19]**	0.82 [0.62, 1.02]**	0.53 [0.28, 0.81]**	0.31 [0.04, 0.58]*	0.50 [0.38, 0.61]**
Math Disorder	1 – 4	192	0.85 [0.53, 1.16]**	1.67 [1.36, 1.98]**	0.82 [0.58, 1.05]**	0.75 [0.56, 0.94]**	1.02 [0.79, 1.25]**
Nonverbal LD	1, 2	39	0.44 [0.13, 0.73]*	1.78 [1.47, 2.09]**	0.99 [0.59, 1.19]**	0.89 [0.60, 1.18]**	1.46 [1.16, 1.76]**
ADHD - Inattentive Type	1 – 4	704	0.88 [0.76, 1.00]**	1.03 [0.86, 1.20]**	0.68 [0.49, 0.87]**	0.63 [0.43, 0.83]**	1.14 [0.98, 1.30]**
ADHD - Combined Type	1 – 4	715	0.99 [0.81, 1.17]**	0.73 [0.64, 0.82]**	1.11 [0.93, 1.29]**	0.69 [0.55, 0.82]**	1.20 [1.00, 1.39]**
Disruptive Disorders[c]	1 – 4	441	0.49 [0.25, 0.73]**	0.41 [0.31, 0.51]**	1.07 [0.83, 1.31]**	0.60 [0.36, 0.85]**	0.46 [0.20, 0.72]**
Pervasive Dev. Disorder[d]	1,2,4	49	0.79 [0.49, 1.08]**	0.98 [0.68, 1.28]**	2.88 [2.61, 3.13]**	2.23 [1.94, 2.52]**	1.52 [1.14, 1.91]**
Mood Disorders[e]	1 – 4	221	0.61 [0.39, 0.83]**	0.75 [0.50, 1.00]**	0.94 [0.59, 1.29]**	1.14 [0.86, 1.42]**	0.93 [0.55, 1.31]**
Anxiety Disorders[f]	1 – 4	364	0.46 [0.24, 0.66]**	0.52 [0.28, 0.76]**	0.35 [0.03, 0.66]*	0.95 [0.72, 1.17]**	0.46 [0.22, 0.69]**
Significant group differences[g]			RD > all others; IT, CT > Anx, DBD	MD > all others; IT > CT, DBD, Anx	PDD > all others; CT > RD, IT, Anx; DBD > RD, Anx	PDD > all others; Mood > CT, RD Anx > RD	MD, NVLD, PDD, IT, CT > RD, DBD, Anx

Note.

[a] the population mean and SD is estimated from the unselected sample of participants in the community sample (see Table 2). d_W = overall effect size based on random effects model (e.g., DerSimonian & Laird, 1986).

[b] 1. Denver clinic, 2. Boulder clinic, 3. twin sample, 4. community sample.

[c] includes diagnoses of oppositional defiant disorder (N = 278), conduct disorder (N = 157), and adjustment disorder with disturbance of conduct (N = 6).

[d] includes diagnoses of aspergers disorder (N = 18), autistic disorder (N = 13), and pervasive developmental disorder, not otherwise specified (N = 18).

[e] includes diagnoses of major depressive disorder (N = 117), dysthymic disorder (N = 55), bipolar disorder (N = 24), and mood disorder not otherwise specified (N = 25).

[f] included diagnoses of generalized anxiety disorder (N = 255), separation anxiety disorder (N = 107), and anxiety disorder, not otherwise specified (N = 2).

[g] RD = reading disorder, MD = math disorder, NVLD = nonverbal learning disorder, IT = ADHD - Inattentive Type, CT = ADHD - Combined Type, Anx = Anxiety Disorder, DBD = disruptive behavior disorder, PDD = pervasive developmental disorder, Mood = mood disorder.

* = P < .01

** = P < .001.