Colorado Learning Difficulties Questionnaire:
Validation of a Parent-Report Screening Measure

Erik G. Willcutt
University of Colorado, Boulder

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

Margaret W. Riddle
University of Denver

Nomita Chhabildas and John C. DeFries
University of Colorado, Boulder

Bruce F. Pennington
University of Denver

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 2 large community samples and 2 samples from clinics that specialize in the assessment of learning disabilities and related disorders (total N = 8,004). Exploratory and confirmatory factor analyses revealed 5 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 4 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 Diagnostic and Statistical Manual of Mental Disorders (4th ed., Text Rev.; DSM–IV; American Psychiatric Association, 2000) provides diagnostic criteria for reading disorder (RD), math disorder (MD), and disorder of written expression. In addition to these DSM–IV categories, other authors described nonverbal 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 (PDDs; e.g., Klin, Volkmar, Sparrow, Cicchetti, & Rourke, 1995; Rourke, 1989).

LDs are associated with a range of negative outcomes and significant public 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 because of 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, Williams, 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, Loeb, & Stouthamer-Loeb, 2003; McGee et al., 2002; Semrud-Clikeman et al., 1992; Trzesniewski, Moffitt, Caspi,
Taylor, & Maughan, 2006; Willcutt et al., 2007; Willcutt & Pennington, 2000a, 2000b).

The high prevalence of LDs and their frequent co-occurrence with other disorders suggests that LD assessment measures should be included systematically 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 Gad 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 LDs and related developmental difficulties. In this article, 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 (EFAs) were completed in each sample, and a subsequent multigroup confirmatory factor analysis (CFA) 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 interrater 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 1 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 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 (see 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 (see 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 interrater 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 children younger than 6 years old because 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 5-point Likert-type 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 because of 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 five 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 interrater 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 article.

Internal structure of the final scale. An initial 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 (see 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 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. Because of 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 chi-squared (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. Interrater 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 1 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 (see 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 (see Table 5). Because of 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 note 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 (Woodcock & Johnson, 1990; Woodcock, McGrew, & Mather, 2001), the Peabody Individual Achievement Test (Dunn & Markwardt, 1970), the Gray Oral Reading Tests (Wiederhold & Bryant, 1993, 2001), and the Wide Range Achievement Test (Jastak & Wilkinson, 1984; Wilkinson, 1993), all of which are widely used standardized measures of academic achievement.

Social functioning. The Behavior Assessment System for Children (BASC; Reynolds & Kamphaus, 1992, 2004) and Achenbach System of Empirically Based Assessment (ASEBA; Achen-
bach & Rescorla, 2001) 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; Rey, 1941), a task which requires the participant to copy a complex figure. A subset

### Table 3
Principal Axis Factor Analyses and Reliability of the Final Pool of CLDQ Items

<table>
<thead>
<tr>
<th>Item</th>
<th>Inter-rater</th>
<th>Test-retest</th>
<th>Primary factor loading in the four samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Difficulty with spelling</td>
<td>.67</td>
<td>.69</td>
<td>.70 (.60, .77)</td>
</tr>
<tr>
<td>2. Difficulty learning letters</td>
<td>.58</td>
<td>.60</td>
<td>.65 (.55, .75)</td>
</tr>
<tr>
<td>3. Difficulty learning phonics</td>
<td>.67</td>
<td>.68</td>
<td>.80 (.75, .87)</td>
</tr>
<tr>
<td>4. Reads slowly</td>
<td>.74</td>
<td>.69</td>
<td>.86 (.80, .91)</td>
</tr>
<tr>
<td>5. Reads below grade level</td>
<td>.75</td>
<td>.75</td>
<td>.87 (.83, .90)</td>
</tr>
<tr>
<td>6. Required extra reading help</td>
<td>.78</td>
<td>.73</td>
<td>.85 (.80, .89)</td>
</tr>
</tbody>
</table>

### Factor 1: Reading
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 2: Social cognition
11. Difficulty making or keeping friends .57 .64 .66 (.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 3: Social anxiety
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 4: Spatial
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)

**Note.** CLDQ = Colorado Learning Difficulties Questionnaire.

### Table 4
Reliability of CLDQ Scores and Correlations Between Scales

<table>
<thead>
<tr>
<th>Composite score</th>
<th>Reliability of composite scores</th>
<th>Correlations between composite scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading</td>
<td>Cronbach’s $\alpha$ M (SD)$^b$</td>
<td>$r_w$ (95% CI) Social cognition $r_w$ (95% CI) Social anxiety $r_w$ (95% CI) Spatial $r_w$ (95% CI)</td>
</tr>
<tr>
<td>Reading</td>
<td>1.79 (0.94) .90 (.88, .93) .83 .81</td>
<td>— — —</td>
</tr>
<tr>
<td>Social cognition</td>
<td>1.54 (0.73) .86 (.83, .88) .53 .71</td>
<td>.28 (.15, .40) — —</td>
</tr>
<tr>
<td>Social anxiety</td>
<td>1.60 (0.75) .82 (.79, .83) .59 .68</td>
<td>.22 (.09, .33) .56 (.52, .59) — —</td>
</tr>
<tr>
<td>Spatial</td>
<td>1.77 (0.92) .85 (.82, .88) .69 .76</td>
<td>.37 (.25, .48) .38 (.32, .44) .33 (.28, .37) — —</td>
</tr>
<tr>
<td>Math</td>
<td>1.73 (0.88) .80 (.76, .82) .63 .73</td>
<td>.32 (.22, .42) .25 (.18, .32) .23 (.18, .28) .47 (.41, .53)</td>
</tr>
</tbody>
</table>

**Note.** CLDQ = Colorado Learning Difficulties Questionnaire; CI = confidence interval.

$^a$ $r_w$ = overall correlation and 95% CI were estimated with random effects models described by DerSimonian & Laird (1986). Total $N = 7,634$. $^b$ The population mean and standard deviation were 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 ($N = 524$).
of the Denver clinic sample completed the Developmental Test of Visual-Motor Integration (DTVMI; Beery & Buktenica, 1997), a standardized measure that requires the participant to copy a series of increasingly complex designs. Finally, the twin study included Primary Mental Abilities (PMA; Thurstone, 1963) 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. Presentation errors of the Denver clinic sample completed the Developmental Test of Visual-Motor Integration (DTVMI; Beery & Buktenica, 1997), a standardized measure that requires the participant to copy a series of increasingly complex designs. Finally, the twin study included Primary Mental Abilities (PMA; Thurstone, 1963) 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 ASEA and 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; Gadow & Sprafkin, 1998), Child Symptom Inventory (CSI; Gadow & Sprafkin, 1998), Social Problems Inventory (ASI), and the BASC. The measures are administered to a subset of cases depending on presenting concerns. The BASC was only included in the standard screening packet in a subset of school districts to comply with requests of district administrators. Administration of the subset of cases was determined by presenting concerns. The BASC was only included in the standard screening packet in a subset of school districts to comply with requests of district administrators.

### Table 5
Measures Used to Evaluate the External Validity of the CLDQ Scales

<table>
<thead>
<tr>
<th>Construct</th>
<th>Denver clinic (Total N = 954)</th>
<th>Boulder clinic (Total N = 179)</th>
<th>Twins (Total N = 1,840)</th>
<th>Community sample (Total N = 5,031)$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single-word reading</td>
<td>WJ LWID 928</td>
<td>WJ-III LWID 177</td>
<td>PIAT read. 1,840</td>
<td>WJ-III LWID 1,564$^a$</td>
</tr>
<tr>
<td>Reading comprehension</td>
<td>GORT-III 847</td>
<td>GORT-IV 161</td>
<td>PIAT reading comp. 1,840</td>
<td>— —</td>
</tr>
<tr>
<td>Math</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calculations</td>
<td>WJ calculations 928</td>
<td>WJ-III calculations 177</td>
<td>WRAT math 1,660</td>
<td>WJ-III calculations 1,564$^a$</td>
</tr>
<tr>
<td>Word problems</td>
<td>WJ applied prob. 867</td>
<td>WJ-III applied prob. 177</td>
<td>PIAT math 1,656</td>
<td>— —</td>
</tr>
<tr>
<td>Social functioning</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social problems</td>
<td>ASEA 642$^b$</td>
<td>BASC-II 178</td>
<td>ASEA 1,815</td>
<td>BASC 2,224$^c$</td>
</tr>
<tr>
<td>Withdrawn behaviors</td>
<td>ASEA 642$^b$</td>
<td>BASC-II 178</td>
<td>ASEA 1,815</td>
<td>BASC 2,224$^c$</td>
</tr>
<tr>
<td>Social skills</td>
<td>— — BASC-II 178</td>
<td>— — BASC 2,224$^c$</td>
<td>— — BASC 2,224$^c$</td>
<td>— —</td>
</tr>
<tr>
<td>Sociometric ratings</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liked by peers</td>
<td>— — No. of friends 172</td>
<td>Dishion (1990) 1,474</td>
<td>Dishion (1990) 5,025</td>
<td>— —</td>
</tr>
<tr>
<td>Spatial functioning</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Block design</td>
<td>WISC-III 803$^b$</td>
<td>WISC-IV 173$^b$</td>
<td>WISC-R 1,840</td>
<td>WISC-III 1,564$^a$</td>
</tr>
<tr>
<td>Other spatial</td>
<td>ROCFT copy 324$^d$</td>
<td>ROCFT copy 121$^d$</td>
<td>ROCFT copy 308$^d$</td>
<td>— —</td>
</tr>
<tr>
<td>DTVMI</td>
<td>478$^d$</td>
<td>— —</td>
<td>PMA spatial 1,840</td>
<td>— —</td>
</tr>
<tr>
<td>Psychopathology</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anxiety</td>
<td>ASEA$^a$ 642$^b$</td>
<td>BASC-II 174</td>
<td>ASEA$^a$ 1,815</td>
<td>BASC 2,224$^c$</td>
</tr>
<tr>
<td>Depression</td>
<td>ASEA$^a$ 642$^b$</td>
<td>BASC-II 174</td>
<td>ASEA$^a$ 1,815</td>
<td>BASC 2,224$^c$</td>
</tr>
<tr>
<td>ADHD</td>
<td>— —</td>
<td>— —</td>
<td>DICA-IV GAD/SAD 1,840</td>
<td>DICA-IV GAD/SAD 716$^f$</td>
</tr>
<tr>
<td>Delinquent behavior</td>
<td>ASEA 642$^b$</td>
<td>BASC-II 175</td>
<td>ASEA 1,815</td>
<td>BASC 2,224$^c$</td>
</tr>
<tr>
<td>PPD</td>
<td>Clinical diagnosis 15</td>
<td>CSI/ASI 175</td>
<td>Exclusion criterion —</td>
<td>Parent report 24</td>
</tr>
</tbody>
</table>

**Note.** CLDQ = Colorado Learning Difficulties Questionnaire; WI-R and WI-III = Woodcock-Johnson Tests of Achievement (WI-R = revised, Woodcock & Johnson, 1990; WI-III = 3rd ed., Woodcock, McGrew, & Mather, 2001); LWID = Letter Word Identification; PIAT = Peabody Individual Achievement Test (Dunn & Markwardt, 1970); GORT = Gray Oral Reading Test (GORT-III = 3rd ed., Wiederhold & Bryant, 1993; GORT-IV = 4th ed., Wiederhold & Bryant, 2001); WRAT = Wide Range Achievement Test (Jastak & Wilkinson, 1984; Wilkinson, 1993); ASEA = peer (Child Behavior Checklist) and teacher (Teacher Report Form) ratings on the Achenbach System of Empirically Based Assessment (Achenbach, 2001); BASC = parent and teacher ratings on the Behavior Assessment System for Children (BASC, Reynolds & Kamphaus, 1992; BASC-II = 2nd ed., Reynolds & Kamphaus, 2004); WISC = Wechsler Intelligence Scale for Children (WISC-R = revised, Wechsler, 1974; WISC-III = 3rd ed., Wechsler, 1991; WISC-IV = 4th ed., Wechsler, 2003); 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); PMA Spatial = Primary Mental Abilities Spatial Relations subtest (Thurstone, 1963); CSI = Child Symptom Inventory (Gadow & Sprafkin, 1997a); ASI = Adolescent Symptom Inventory (Gadow & Sprafkin, 1998); GAD = generalized anxiety disorder; DICA-IV = DSM–IV Diagnostic Interview for Children and Adolescents (Reich, Welner, & Herjanic, 1997); SAD = social anxiety disorder; MDD = major depressive disorder; ADHD = attention-deficit/hyperactivity disorder; DBRS = Disruptive Behavior Rating Scale (Barkley & Murphy, 1998); PDD = pervasive developmental disorder. 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.

$^f$ 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.

$^g$ Administered previously as part of the standard test battery, then discontinued.

$^h$ CSI and ASI ratings of PDD symptoms were used for correlational analyses, and PDD diagnoses were made based on all available clinical information.
Gadow & Sprafkin, 1997a), or DSM–IV Diagnostic Interview for Children and Adolescents (DICA-IV; Reich, Welner, & Herjanic, 1997).

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 3 standard deviations below the mean and more than 0.5 standard deviations 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 ($rs = .06–.13$). 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 nonindependence 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 (Muthén & Muthén, 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 $d$ 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 (e.g., 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 with 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 (see Table 6), and ratings of all clinical groups were significantly higher than the estimated population mean on all CLDQ scales (see 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 (see Table 6). Similarly, the effect size of the difference between the RD group and the estimated population mean was large ($d_w = 1.81$; see 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 clinical group (see 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.

Math Scale

Measures of math achievement were more highly correlated with the CLDQ Math scale than the other four CLDQ scales (see 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 did any other clinical group (see Table 7), and in the group with MD, the effect size on the Reading 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 (see Table 6). Group comparisons indicated that groups with a PDD scored significantly higher on the Social Cognition scale than did any of the other clinical groups and were more impaired on the scale than any of the other CLDQ scales (see 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 (see 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 disorders, CLDQ Social Anxiety scores were more strongly associated with social isolation, withdrawn behaviors, and anxiety disorders (see 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

### Table 6

**Correlations Between CLDQ Scales and Measures Used for External Validation**

<table>
<thead>
<tr>
<th>Measure of related constructs</th>
<th>Samples</th>
<th>Total N</th>
<th>Reading $r_w$ (95% CI)</th>
<th>Math $r_w$ (95% CI)</th>
<th>Social Cognition $r_w$ (95% CI)</th>
<th>Social Anxiety $r_w$ (95% CI)</th>
<th>Spatial $r_w$ (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic achievement</td>
<td>1–4</td>
<td>4,518</td>
<td>.64 (.60, .68)**</td>
<td>.19 (.10, .28)**</td>
<td>.16 (.10, .22)**</td>
<td>.07 (−.03, .16)**</td>
<td>.17 (.10, .23)**</td>
</tr>
<tr>
<td>Math composite</td>
<td>1–4</td>
<td>4,518</td>
<td>.35 (.28, .40)**</td>
<td>.19 (.10, .28)**</td>
<td>.16 (.10, .22)**</td>
<td>.07 (−.03, .16)**</td>
<td>.17 (.10, .23)**</td>
</tr>
<tr>
<td>Social functioning</td>
<td>1–4</td>
<td>7,686</td>
<td>.20 (.16, .24)**</td>
<td>.25 (.22, .28)**</td>
<td>.44 (.35, .53)**</td>
<td>.28 (.26, .31)**</td>
<td>.23 (.19, .27)**</td>
</tr>
<tr>
<td>Social rejection composite</td>
<td>1–4</td>
<td>7,686</td>
<td>.17 (.12, .21)**</td>
<td>.22 (.17, .27)**</td>
<td>.27 (.24, .30)**</td>
<td>.46 (.36, .56)**</td>
<td>.19 (.13, .25)**</td>
</tr>
<tr>
<td>Social isolation composite</td>
<td>1–4</td>
<td>6,436</td>
<td>.22 (.20, .25)**</td>
<td>.12 (.08, .16)**</td>
<td>.41 (.32, .49)**</td>
<td>.26 (.20, .31)**</td>
<td>.24 (.21, .26)**</td>
</tr>
<tr>
<td>Social strengths composite</td>
<td>2–4</td>
<td>6,436</td>
<td>.16 (.11, .21)**</td>
<td>.35 (.30, .40)**</td>
<td>.10 (−.07, .13)**</td>
<td>.07 (−.04, .10)**</td>
<td>.30 (.27, .33)**</td>
</tr>
<tr>
<td>Spatial functioning</td>
<td>1–4</td>
<td>4,511</td>
<td>.16 (.11, .21)**</td>
<td>.35 (.30, .40)**</td>
<td>.10 (−.07, .13)**</td>
<td>.07 (−.04, .10)**</td>
<td>.30 (.27, .33)**</td>
</tr>
<tr>
<td>Psychopathology</td>
<td>1–4</td>
<td>7,686</td>
<td>.13 (.11, .15)**</td>
<td>.14 (.11, .17)**</td>
<td>.23 (.16, .30)**</td>
<td>.36 (.32, .40)**</td>
<td>.12 (.07, .16)**</td>
</tr>
<tr>
<td>Anxiety composite</td>
<td>1–4</td>
<td>7,686</td>
<td>.13 (.11, .15)**</td>
<td>.14 (.11, .17)**</td>
<td>.23 (.16, .30)**</td>
<td>.36 (.32, .40)**</td>
<td>.12 (.07, .16)**</td>
</tr>
<tr>
<td>Depression symptoms</td>
<td>2–4</td>
<td>3,744</td>
<td>.20 (.17, .22)**</td>
<td>.19 (.15, .23)**</td>
<td>.28 (.24, .31)**</td>
<td>.32 (.26, .37)**</td>
<td>.22 (.19, .25)**</td>
</tr>
<tr>
<td>Externalizing behaviors</td>
<td>1–4</td>
<td>7,686</td>
<td>.11 (.07, .15)**</td>
<td>.09 (.01, .17)**</td>
<td>.43 (.39, .47)**</td>
<td>.23 (.19, .27)**</td>
<td>.12 (.08, .16)**</td>
</tr>
<tr>
<td>Inattentiveness</td>
<td>1–4</td>
<td>7,815</td>
<td>.46 (.40, .51)**</td>
<td>.32 (.22, .42)**</td>
<td>.52 (.41, .63)**</td>
<td>.32 (.27, .38)**</td>
<td>.56 (.50, .62)**</td>
</tr>
<tr>
<td>Hyperactivity - impulsivity</td>
<td>1–4</td>
<td>7,815</td>
<td>.28 (.21, .34)**</td>
<td>.16 (.10, .22)**</td>
<td>.47 (.34, .60)**</td>
<td>.20 (.14, .26)**</td>
<td>.36 (.29, .43)**</td>
</tr>
<tr>
<td>PDD symptoms</td>
<td>2</td>
<td>179</td>
<td>.08 (−.07, .22)</td>
<td>.24 (.11, .37)**</td>
<td>.64 (.54, .74)**</td>
<td>.41 (.27, .54)**</td>
<td>.16 (.01, .30)**</td>
</tr>
</tbody>
</table>

**Note.** CLDQ = Colorado Learning Difficulties Questionnaire; CI = confidence interval; PDD = pervasive developmental disorder.

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 (DerSimonian and Laird, 1986). b (1) Denver clinic, (2) Boulder clinic, (3) twin sample, (4) community sample.

* p < .01. ** p < .001.
Cognition, or Social Anxiety scales (see Table 6), but this association was similar in magnitude to the correlation for Social Anxiety disorders.  

Note. CLDQ = ; CI = confidence interval; LD = learning disorder; ADHD = attention-deficit/hyperactivity disorder; PDD = pervasive developmental disorder; MD = math disorder; Mood = mood disorder; NVLD = nonverbal learning disorder.

The population mean and standard deviation are estimated from the selected sample of participants in the community sample. $d_\mu$ = overall effect size based on random effects model (e.g., DerSimonian & Laird, 1986). $^a$(1) Denver clinic, (2) Boulder clinic, (3) twin sample, (4) community sample. $^b$ Includes diagnoses of oppositional defiant disorder ($n = 278$), conduct disorder ($n = 157$), and adjustment disorder with disturbance of conduct ($n = 6$). $^c$ Includes diagnoses of Asperger’s disorder ($n = 18$), autistic disorder ($n = 13$), and PDD not otherwise specified ($n = 18$). $^d$ Includes diagnoses of major depressive disorder ($n = 117$), dysthymic disorder ($n = 55$), bipolar disorder ($n = 24$), and mood disorder not otherwise specified ($n = 25$). $^e$ Includes diagnoses of generalized anxiety disorder ($n = 255$), separation anxiety disorder ($n = 107$), and anxiety disorder not otherwise specified ($n = 2$).

$p < .01$. $^*^* p < .001$.

Cognition, or Social Anxiety scales (see 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 (see Table 7). However, consistent with the results of the dimension 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 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. EFAs of the CLDQ identified five factors in all four samples, and CFAIs 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 interrater and test–retest reliability. Convergent evidence for the CLDQ Reading scale is provided by significant correlations with standardized measures of reading achievement (overall $r = .61$). 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, 2010). 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 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 clinical 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 clinical 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 ($n = 589$) with 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 ($d = 0.19$), 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 ($r_s = .27–.32$).

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 because of 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 important, 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 sample ($d_w = 1.64$). 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 on the CLDQ Reading scale was higher, but no 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). Because of the relatively small size of the initial item pool (46 items) and the exclusion of over half of the 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, Pick, & 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 because of 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, compared with 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 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.

References


