Published in final edited form as:

Clin Psychol Rev. 2008 July ; 28(6): 1059–1070. doi:10.1016/j.cpr.2008.03.001.

Categorical and dimensional approaches to the measurement of disruptive behavior in the preschool years: A meta-analysis

Angela D. Moreland* and Jean E. Dumas¹

Department of Psychological Sciences, Purdue University, 703 Third Street, West Lafayette, IN, 47907, USA

Abstract

Preschool disruptive behavior problems were investigated in a meta-analysis of 26 studies using categorical and/or dimensional approaches to assessment. The review sought to distinguish early disruptiveness from normative preschool conduct by showing that, irrespective of assessment methodology: (a) disruptiveness can be adequately measured in the preschool years; (b) early disruptiveness is stable over time; and (c) disruptive children referred for clinical services in the preschool years can be distinguished from non-referred peers. Results indicated that: categorical and dimensional approaches to measurement of early disruptiveness provide comparable data (effect size $d = 2.29$); both approaches yield comparable estimates of the stability of preschool disruptive behavior over time (categorical approach: $d = 1.15$; dimensional approach: $d = 0.84$); and both approaches discriminate between referred and non-referred preschoolers ($d = 1.05$ and $d = .95$). Limitations of the existing literature and of this analysis are discussed, as are suggestions for future research.

Keywords

Early disruptive behavior; Preschool; Meta-analysis; Categorical approach; Dimensional approach

1. Introduction

Two to six percent of children in the general population are diagnosed with clinically significant levels of disruptive behavior problems. These problems are a source of concern for schools, families, and public health, and are the most frequent reason for referral to child mental health clinics (APA, 2000; Keenan & Wakschlag, 2000). Behavior problems can begin in the preschool years, in many cases as early as 3 to 4 years of age (Campbell, Ewing, Breaux, & Szumowski, 1986; Keenan, Shaw, Delliquadri, Giovannelli, & Walsh, 1998; Loeber, 1990). When they do, prognosis is guarded, as many “early starters” remain disruptive throughout childhood and beyond (Earls, 1980; White, Moffitt, Earls, & Robbins, 1990; Zahn-Waxler, Ianotti, Cummings, & Denham, 1990). For example, Campbell (1990) reported that approximately 50% of preschoolers displaying disruptive behavior problems

*Corresponding author. Tel.: +1 765 494 6991; fax: +1 765 496 2670. moreland@psych.purdue.edu (A.D. Moreland), jed@psych.purdue.edu (J.E. Dumas).

¹Present address: Université de Genève, FAPSE - Section de psychologie, Bd du Pont d'Arve 40, CH-1211 Genève 4, Switzerland.

continued to show clinically significant levels of difficulties when they reached school age, with two-thirds of the overall preschoolers meeting criteria for a disruptive behavior disorder by age 9 (i.e., Attention Deficit/Hyperactivity Disorder or ADHD, Oppositional Defiant Disorder or ODD). Comparable findings have been reported by Egeland, Kalkoske, Gottesman, and Erickson (1990) and Loeber (1982). In sum, disruptive behavior problems are associated with chronic, often life-long challenges, especially when they begin in the preschool years.

The study of preschool disruptive behavior problems is complicated by the fact that two conceptual approaches – categorical and dimensional – are commonly used to assess such problems and do not always yield comparable data (Eyberg, Schuhmann, & Rey, 1998; Keenan & Wakschlag, 2000; Waldman, Lilienfeld, & Lahey, 1995). A categorical approach to assessment relies on diagnostic criteria to determine the presence or absence of disruptive or other abnormal behaviors (e.g., Diagnostic and Statistical Manual of Mental Disorders or DSM-IV, APA, 2000), whereas a dimensional approach places such behaviors on a continuum of frequency and/or severity (e.g., Child Behavior Checklist or CBCL, Achenbach & Edelbrock, 1983; Lavigne et al., 1996). Both approaches have notable merits and drawbacks, and are associated with a long tradition of research, such as Paul Meehl's (2001) taxometric work and Jacob Cohen's (1983) work on loss of explanatory power when continuous measures are dichotomized (see Beauchaine, 2003, for a detailed review and discussion).

1.1. Comparing categorical and dimensional approaches to early disruptive behavior

Research on disruptive behavior in children and adolescents has focused, among other things, on a comparison of categorical and dimensional approaches to assessment in an attempt to identify their similarities and differences. Studies show that both approaches yield overlapping data, but their results are often difficult to compare, mainly because they rely on different measurement tools and analytical methods (Scholte, Van Berckelaer-Onnes, & Van Der Ploeg, 2002; Teagarden & Burns, 1999).

To date, only two studies have focused on a comparison of categorical and dimensional approaches to the assessment of preschool disruptive behavior problems (Pelletier, Collett, Gimpel, & Crowley, 2006; Sprafkin, Volpe, Gadow, Nolan, & Kelly, 2002). These studies relied on different instruments to measure these problems — specifically, on two categorical and on two dimensional rating scales. Pelletier et al. (2006) assessed preschoolers using a categorical instrument, the Disruptive Behavior Disorders Rating Scale (DBDRS), and a dimensional instrument, the School Situations Questionnaire (SSQ), and found also that data obtained with each approach were significantly correlated in the measurement of Attention Deficit/Hyperactivity Disorder (ADHD) ($r=.77, p<.01$) and Oppositional Defiant Disorder (ODD) ($r=.69, p<.01$). Sprafkin et al. (2002) used the Early Childhood Inventory-4 (ECI-4) as a categorical rating scale, and the Child Behavior Checklist (CBCL) as a dimensional scale, for purposes of comparison as well. They found that scores obtained with these two approaches were again highly correlated, both in the measurement of ADHD ($r=.70, p<.001$) and of ODD ($r=.81, p<.001$). These two studies strongly suggest that early disruptiveness can be reliably measured in the preschool years with both categorical and

dimensional instruments. However, pooling their samples and statistics, as meta-analytical procedures allow, can provide a quantitative estimate of the extent to which the two approaches yield comparable results, even though those results are based on a variety of research instruments.

1.2. Estimating the stability of early disruptive behavior

Although a number of studies using both categorical and dimensional approaches has examined stability of early disruptive behavior over time (Campbell, 1994; Campbell & Ewing, 1990; Crawford & Lee, 1991; Gadow, Sprafkin, & Nolan, 2001; Olson & Brodfield, 1991), comparisons of their findings are again limited by their use of a variety of assessment instruments.

In general, categorical and dimensional studies show that early disruptive behavior problems tend to be stable. Considering categorical evidence first, Gadow et al. (2001) found that preschoolers who had been rated by both parents and teachers as having symptoms of disruptiveness on the ECI-4 continued to display such symptoms eight months later on the same instrument ($r=.46, p<.001$). Similarly, Lahey, Pelham, Loney, Lee, and Willcutt (2005) asked parents and teachers to rate preschoolers on symptoms of disruptive behavior using the Diagnostic Interview Schedule for Children (DISC) and the DBDRS at baseline and at a one-year follow-up. Results showed that disruptive behavior was relatively stable over the study period ($r=.23, p<.001$). Finally, Pierce, Ewing, and Campbell (1999) demonstrated the stability of disruptive behavior problems over a ten-year period. That study assessed such problems at age 3 using the Swanson, Nelson, and Pelham Scale (SNAP) completed by parents, at age 9 using the SNAP again, and at age 13 using a semi-structured interview. Results showed that 67% of children identified as having behavior problems at age 3 met criteria for an externalizing disorder at age 9, and that over half of them continued to engage in disruptive behavior at age 13. Turning to dimensional studies, Crawford and Lee (1991) assessed preschoolers twice with the CBCL over a 6-week period and found that ratings of externalizing problems held over time. Similarly, Campbell (1994) used the same instrument to obtain parent ratings of children at age 3 and 6. Results showed again that children with elevated ratings of externalizing problems at age 3 tended to have comparably elevated ratings three years later.

Although categorical and dimensional data point to the stability of early disruptive behavior problems, comparisons across studies are made difficult for several reasons. First, studies are not uniform in their definition of these problems. For example, Gadow et al. (2001) defined early disruptiveness in terms of separate categorical ratings of ADHD and ODD, Campbell (1994) in terms of categorical ratings of ADHD only, and Campbell and Ewing (1990) in terms of categorical ratings of ADHD and dimensional ratings of externalizing problems. Second, studies vary greatly in the time span used to establish stability, intervals between baseline and follow-up ranging from 6 weeks to 10 years. Finally, studies vary considerably in the statistics they use to establish stability of early disruptiveness, some reporting chi-square values (e.g., Egeland et al., 1990; Pierce et al., 1999), others correlations (e.g., Crawford & Lee, 1991; Gadow et al., 2001), and still others percentages of children who continued to display disruptive behavior at follow-up (e.g., Campbell, 1994; Campbell et al.,

1986). Converting these different statistics into effect sizes, as is done in meta-analysis, can greatly facilitate the interpretation of results across studies by providing quantitative estimates of stability for both categorical and dimensional approaches.

1.3. Comparing disruptive behavior in referred and non-referred preschoolers

Several studies have also examined whether categorical and/or dimensional approaches are capable of discriminating between preschoolers who are referred for clinical services because of disruptive behavior problems and their peers who are not. In general, both approaches can do so successfully. For example, researchers have used a variety of research instruments in a categorical manner to show that referred children can be distinguished from non-referred peers, including the SNAP (Campbell, 1994), the ECI- 4 (Gadow et al., 2001), the Preschool Behavior Questionnaire (PBQ) (Pierce et al., 1999), and the Kiddie-Disruptive Behavior Disorders Scale (K-DBDS) (Keenan et al., 2007). Only three studies exist in which researchers chose a dimensional approach to address the same question. Relying in each case on the CBCL, their results show that referred preschoolers have significantly higher scores ($p < .001$) than their non-referred peers on all subscales of the instrument (Achenbach, Edelbrock, & Howell, 1987; Campbell & Ewing, 1990; Heller, Baker, Henker, & Hinshaw, 1996). Taken together, categorical and dimensional studies lead to similar conclusions when referred and non-referred children are compared, even though they often differ in their definitions of early disruptiveness and, in many cases, suffer from the same limitations we mentioned above with respect to stability.

1.4. Aims of the present study

To integrate the existing literature on preschool disruptive behavior problems, a meta-analytical review of studies using categorical and/or dimensional approaches to the assessment of early disruptiveness was conducted. The review sought to distinguish early disruptiveness from normative preschool conduct by showing that, irrespective of assessment methodology: (a) disruptiveness can be adequately measured in the preschool years; (b) early disruptiveness is stable over time; and (c) disruptive children referred for clinical services in the preschool years can be distinguished from their non-referred peers.

2. Methods

2.1. Literature search

Studies included in the review were obtained from several sources: (a) searches of PsycINFO and MEDLINE (keywords: disruptive behavior, preschool, categorical, dimensional), (b) manual searches of the journals most likely to publish studies on preschool disruptive behavior, as determined through the literature search², (c) letters sent to 13 researchers who have published articles on the assessment of early disruptiveness, requesting copies of unpublished manuscripts and/or articles currently in press, to which 9 researchers responded, and (d) inspection of reference lists of all pertinent studies. The literature search

²Development and Psychopathology, Journal of Abnormal Child Psychology, Journal of Child Psychology and Psychiatry, Journal of Clinical Child Psychology, Journal of the American Academy of Child and Adolescent Psychiatry.

included studies published from 1986 to 2006, as research on disruptive behavior in preschoolers has primarily been conducted over the past 20 years.

2.2. Inclusion criteria

To be included in the review, studies had to meet the following criteria:

1. The study was written in English.
2. The study was conducted on children aged 2–7 years, which is the common age range used to define preschoolers in the literature. The literature search yielded six studies in which the children's age range was broader (e.g., 2–19 years); they were excluded from the review.
3. Sample size was at least 25, to exclude single case studies from the analyses and ensure adequate power.
4. Informants were parents and/or teachers of the child participants in the study.
5. Studies included one or more categorical and/or dimensional measures of preschool disruptive behavior problems and addressed at least one of the three research questions.
6. The study compared two or more groups to enable calculation of effect sizes (e.g., referred versus non-referred children).
7. Each study relied on a unique sample, to ensure that the same children were not included in more than one individual study used in the review (e.g., when a research team published more than one study on the same sample, the first published study was used).

The literature search yielded a total of 26 published or in press studies. Although the search identified additional studies, they were excluded because they did not meet the inclusion criteria (e.g., Gillion, Shaw, Beck, Schonberg, & Lukon, 2002; Vondra, Shaw, Swearingen, Cohen, & Owens, 2001) or did not provide necessary statistics for analyses (e.g., Lavigne et al., 1998; Lavigne et al., 1998; Shaw, Owens, Giovannelli, & Winslow, 2001).

2.3. Coding procedures

The senior author coded the 26 studies using a standardized coding sheet developed for the purpose of this review. Studies were coded for: (1) the research question(s) they addressed; (2) the approach they adopted (categorical, dimensional, or both); (3) sample n 's, means, standard deviations, and percentages for each group; (4) available test statistics, including t , χ^2 , and F values; (5) demographic information, including child gender, age, and ethnicity, and parent age, marital status, income level, and education (when reported); (6) sample type (e.g., referred, community, school); and (7) study setting (e.g., clinic, school). Five of the studies (19%) were independently coded by an additional coder for purposes of reliability. Results showed that inter-rater reliability was satisfactory ($\kappa=.79$).

2.4. Calculation of effect sizes

To address each research question, two groups were identified in each relevant study and effect sizes were calculated to determine differences between groups. Hedge's g was used in analyses of individual effect sizes, and Cohen's d in analyses of overall cumulative effects (i.e., to average several Hedge's g values). In addition to Hedge's g and Cohen's d values, r values are reported in Table 2.

The DSTAT computer program (Johnson, 1989) was used to calculate effect sizes using Hedge's g :

$$g = \frac{M_1 - M_2}{\sigma_{\text{pooled}}}$$

where

$$\sigma_{\text{pooled}} = \sqrt{[(\sigma_1^2 + \sigma_2^2)/2]}.$$

DSTAT takes sample size into account and corrects for method and cross-sectional variance. Hedge's g was calculated from available test statistics, such as t , F , or p values, for each individual study. When there were more than one way of calculating effect size (because a study reported multiple test statistics, such as an F -test as well as means and standard deviations), all effect sizes were calculated and averaged, as recommended by Hedges and Olkin (1985). For categorical studies that measured two types of disruptive behavior (i.e., ADHD and ODD), effect sizes were calculated for each type and then averaged also. This enabled analyses to be conducted for each type of disruptive behavior, as well as for an overall estimate of disruptiveness.

Once all effect sizes of relevance to each research question had been obtained, they were averaged into composite effect sizes (Cohen's d), with a 95% confidence interval drawn around the mean. Composite effect sizes weigh each g by the reciprocal of its variance, thus giving greater weight to studies with larger sample sizes (Hedges & Olkin, 1985). Homogeneity of the g 's was also examined to determine whether studies could be adequately described with a single estimate (Hedges & Olkin, 1985). When effect sizes were heterogeneous, model testing and outlier diagnosis was conducted to explain variability among effect sizes and/or restore homogeneity, using Hedges and Olkin's (1985) *post hoc* statistical procedures. Outlier studies were identified within each research question.

In line with a long-standing practice in this area, effect sizes that fell below 0.20 were described as small, around 0.50 as medium, and above 0.80 as large (Cohen, 1988). Rosenthal's fail-safe N was calculated along with each effect size in an attempt to address the influence that publication bias might have had on the findings (Wolf, 1986; Wortman, 1994). This statistic calculates the number of studies resulting in null results that would be necessary to make significant effect sizes non-significant ($p < .05$) (Rosenthal, 1979).

3. Results

The 26 studies selected for purposes of this review included a total of 4,536 preschoolers between the ages of 2.00 and 6.10 years ($M = 3.86$).³ Most of them were boys (67%) of Caucasian ethnicity (68%). However, these demographic statistics are estimates only, as 11 of the studies did not report data on child gender and/or ethnicity. Furthermore, most of the studies did not contain information on parental age, gender or ethnicity. See Table 1 for further descriptive information about the sample. As some of the 26 studies addressed more than one of the research questions, data from those studies were used more than once to answer those questions.

3.1. Comparing categorical and dimensional approaches to early disruptive behavior

Two studies compared categorical and dimensional approaches to the measurement of early disruptiveness (see Table 2). Effect size calculations for this research question were based on means, standard deviations, Pearson r correlations, and/or t statistics, with larger effect sizes representing greater similarity in the data obtained with each approach. In other words, when comparisons of categorical and dimensional measures produced large effect sizes, the measures were closely associated and yielded comparable information about early disruptive behavior.

Individual effect sizes for the two studies were large (see Table 2), as was their weighted mean effect size ($d = 2.21$, $N = 388$, 95% confidence interval [CI] = 2.08 to 2.34). One outlier was identified in the analyses, but excluding that study did not change the size of the effect. Separate calculations for ADHD and ODD produced comparable effect sizes ($d = 2.17$, $N = 388$, CI = 1.99 to 2.35 and $d = 2.25$, $N = 388$, CI = 2.07 to 2.43, respectively). Rosenthal's fail-safe N indicated that 237 non-significant studies would be necessary to make the weighted mean effect size non-significant.

Overall, both individual and composite effect sizes show that the relation between categorical and dimensional measures of early disruptiveness is highly significant. Practically, this means that preschoolers identified as exhibiting disruptive behavior with categorical measures have a 99% likelihood of being similarly identified with dimensional measures, and thus that early disruptive behavior problems can be successfully identified with either approach.

3.2. Estimating the stability of early disruptive behavior

Twenty-one studies reported data on the stability of early disruptive behavior using categorical or dimensional approaches to measurement, with one study reporting statistics for both approaches separately. Three studies were excluded because they used a categorical approach at baseline and a dimensional one at follow-up, making it impossible to characterize them within a single approach. Each of the remaining studies collected categorical or dimensional ratings of disruptive behavior at baseline and follow-up, with follow-ups ranging from 6 months to 5 years. Effect size calculations were based on means,

³Reported statistics are based only on the studies which provided demographic information on the children.

standard deviations, and Pearson r correlations, with larger effect sizes representing closer associations (i.e., greater stability) at the two time points.

For the six studies that relied on a categorical approach to measure disruptive behavior over time, individual effect sizes were calculated for each type of behavior and then averaged for that study. The six studies yielded individual effect sizes that ranged from medium to large (see Table 2). Their weighted mean effect size was large ($d = 1.15$, $N = 395$, $CI = .99$ to 1.30). Three outliers were identified but the mean effect size remained large when the three studies were excluded from the analyses ($d = 1.12$, $N = 180$, $CI = .89$ to 1.34). Separate effect sizes were calculated for the five studies that measured ADHD and the two that measured ODD separately. Large effect sizes were found for each type of behavior problems (ADHD: $d = .78$, $N = 294$, $CI = .61$ to $.95$; ODD: $d = 1.60$, $N = 193$, $CI = 1.37$ to 1.83). Rosenthal's fail-safe N indicated that 72 studies with non-significant results would be necessary to make the weighted mean effect size non-significant.

Fifteen studies used a dimensional measure, the CBCL or the PA-SEC (drawn from the Dutch version of the CBCL), to assess the stability of early disruptive behavior problems. Eleven of them yielded large effect sizes, one a medium effect size, and the remaining three a small effect size (see Table 2). The weighted mean effect size for these fifteen studies was large ($d = 0.84$, $N = 1329$, $CI = 0.76$ to 0.92). Homogeneity tests identified nine outliers, which included the four studies yielding small and medium effect sizes. The mean effect size became larger with these nine studies excluded from the analyses ($d = 1.52$, $N = 304$, $CI = 1.34$ to 1.70). Rosenthal's fail-safe N showed that 107 studies containing null results would be necessary to make the weighted mean effect size non-significant.

Overall, both individual and composite effect sizes show that early disruptiveness tends to be stable over time, whether it is measured categorically or dimensionally. Practically, this means that preschoolers displaying disruptive behavior problems at baseline are 86% more likely to display similar problems at follow-up than preschoolers who are not disruptive at baseline when assessed with categorical measures. When preschoolers are assessed with dimensional measures, the corresponding figure is 96%.

3.3. Comparing disruptive behavior in referred and non-referred preschoolers

Fourteen studies measured disruptive behavior in referred and non-referred preschoolers, 11 with a categorical approach and three with a dimensional approach. Effect sizes were calculated using means, standard deviations, F values, and χ^2 statistics, in which significant values represented significant differences between referred and non-referred preschoolers. This means that, at the meta-analytical level, larger effect sizes represent larger differences between referred and non-referred preschoolers.

The individual effect sizes of the 11 studies using a categorical approach ranged from small to large (see Table 2). As three of these studies measured ADHD and ODD separately, average effect sizes were calculated for each study and then used in the composite effect size analyses. The weighted mean effect size of the 11 categorical studies was large ($d = 1.04$, $N = 1685$, $CI = .94$ to 1.15). Five outliers were identified through homogeneity analyses, but the effect size remained large when these studies were excluded ($N = 892$, $d = .94$, $CI = .79$

to 1.08). Rosenthal's fail-safe N indicated that 127 studies would be necessary to reduce the effect size to a non-significant value. Thus, results indicate that preschoolers referred for disruptive behavior can be successfully distinguished from preschoolers who were not when using a categorical approach to assessment. Specifically, when assessing groups of referred and non-referred preschoolers, a significant difference was found between the groups 83% of the time.

Separate effect sizes were calculated for six studies that measured ADHD and nine studies that measured ODD (see Table 2). Studies that assessed ADHD yielded a large effect size ($d = 1.07$, $N = 1173$, $CI = .94$ to 1.20). Four outliers were identified but the effect remained large when these studies were excluded ($N = 104$, $d = 1.27$, $CI = .83$ to 1.70). Studies that measured ODD ($N = 1343$) initially showed a medium effect size of $d = .70$ ($CI = .58$ to $.81$). However, that effect became slightly larger once three identified outliers were excluded ($N = 1021$, $d = .75$, $CI = .62$ to $.88$). Rosenthal's fail-safe N results showed that 94 null studies measuring ADHD and 43 measuring ODD would be necessary to make these effects non-significant. In sum, when studies measured ADHD and ODD separately, a categorical approach yielded medium to large differences between referred and non-referred children. Specifically, children referred for ADHD were found to differ from non-referred children 90% of the time, and 77% of the time for children referred for ODD.

The three studies that used a dimensional approach to distinguish referred from non-referred preschoolers all yielded large individual effect sizes (see Table 2). Their composite weighted mean effect size was large as well ($N = 272$, $d = .95$, $CI = .69$ to 1.21). Rosenthal's fail-safe N indicated that 16 studies would be necessary to reduce the result to non-significance. Practically, the large difference between the referred and non-referred groups indicated that, when assessed with a dimensional approach, the groups differed from one another 83% of the time.

4. Discussion

The purpose of this meta-analysis was to compare categorical and dimensional approaches to assessment of preschool disruptive behavior problems in order to determine the extent to which: (a) disruptiveness can be adequately measured in the preschool years; (b) early disruptiveness is stable over time; and (c) disruptive preschoolers referred for clinical services can be distinguished from their non-referred peers. Results of a review of 26 studies published between 1986 and 2006 indicated that: categorical and dimensional approaches provide comparable data when assessing early disruptiveness; both approaches yield comparable estimates of the stability of preschool disruptive behavior over time; and both approaches discriminate between referred and non-referred preschoolers. We will consider each of these findings in turn.

The first purpose of the review was to compare how well categorical and dimensional approaches measured preschool disruptive behavior problems. Results showed that data obtained with both approaches corresponded closely, at the individual study level and when studies were aggregated to obtain a composite estimate of effect. Although only two studies were available to address this issue, their individual effect sizes were large in spite of

important methodological differences, as was their weighted mean effect size. The finding that categorical and dimensional approaches provide similar pictures of early disruptiveness is consistent with previous literature showing that these approaches are complementary and can both be employed in the assessment of young children (Arend, Lavigne, Rosenbaum, Binns, & Christoffel, 1996) — whether in research or clinical practice.

The second research question focused on comparing categorical and dimensional measurements of the stability of disruptive behavior problems in the preschool years. Results showed that, whether assessed categorically or dimensionally, these problems were stable, in many cases over extended periods of time — both at the individual study level and when studies were aggregated to obtain composite estimates of effect. Individual effects ranged from small to large, with 80% of studies yielding large effects. Tests of homogeneity showed that two of the three studies reporting small or medium effects were outliers, and that results did not change when they were removed from the computation of a mean effect size. In other words, with or without those studies, mean effect sizes demonstrating the stability of early disruptiveness were large for both categorical and dimensional approaches, without much difference between them. Support for the stability of early disruptive behavior problems is consistent with a large body of research showing that “early starters” on a developmental trajectory of antisocial conduct face cumulative challenges that, in some cases, persist into adolescence and beyond (e.g., Dodge, Coie, & Lynam, 2006). This finding points also to the importance of prevention and early intervention programs aimed at parents of preschool children, to divert this trajectory before it becomes stable (Dumas, Nissley-Tsiopinis, & Moreland, 2006; Kazdin, 2005; Keenan & Wakschlag, 2000).

The third purpose of this review was to investigate whether categorical and dimensional approaches distinguished between preschoolers who were referred for disruptive behavior problems and preschoolers who were not. Individual effects ranged from small to large, with 73% of studies yielding large effects. Tests of homogeneity showed that the three studies reporting small or medium effects were outliers and that results did not change when they were removed from the computation of a mean effect size. In other words, with or without those studies, categorical and dimensional approaches yielded large mean effect sizes, thereby showing that preschoolers who were referred for clinical services because they were disruptive could be adequately distinguished from their non-referred peers. This is in line with studies reporting that children who are referred for services for disruptiveness often differ from their peers from a young age, both because their oppositional and defiant conduct alarms their caregivers and because it jeopardizes their own development (e.g., Dumas, 1996).

Taken together, results of this meta-analysis confirm that, whether measured categorically or dimensionally, disruptive behavior problems can appear as early as age 3 and that they can be distinguished from the normative opposition and defiance that young children often display. Besides their public health implications for prevention and early intervention already mentioned, these findings suggest that longitudinal studies focusing on the development of behavioral and emotional problems in children and youth may be most informative if they begin early — earlier than a number of important projects that began data

collection in the school-age years (e.g., Braswell, August, Bloomquist, & Realmuto, 1997; Masse & Tremblay, 1999).

4.1. Limitations and suggestions for future research

Important limitations caution against over-interpreting the findings. First, the current study was primarily methodological in nature. Relying on existing literature, it compared categorical and dimensional approaches to the measurement of disruptiveness in the preschool years; it did not focus on identifying the core behaviors that are characteristic of early disruptiveness or on tracing the developmental trajectories of those behaviors — two topics that have drawn considerable research interest in their own right (e.g., Dodge et al., 2006).

Second, although this meta-analysis shows that categorical and dimensional approaches provide complementary data on early disruptiveness, each approach reflects a small set of assessment tools. As we mentioned earlier, commonly used categorical measures include the SNAP (Campbell, 1994), the ECI-4 (Gadow et al., 2001), the Preschool Behavior Questionnaire (PBQ) (Pierce et al., 1999), and the Kiddie-Disruptive Behavior Disorders Scale (K-DBDS) (Keenan et al., in press). Their distinct names suggest that these measures assess disruptiveness in unique ways. However, close examination shows that their content overlaps considerably, as each measure reflects the diagnostic criteria for disruptive behavior (esp. ADHD and ODD) of the DSM-IV (or earlier editions). An even smaller set of measures are used to assess disruptiveness dimensionally. The CBCL dominates this area of research to such an extent that no other rating scale is as widely used and commonly accepted. In other words, results confirm the adequacy of both approaches in the assessment of early disruptive behavior problems, but point to the narrow manner in which “categorical” and “dimensional” are defined empirically.

Third, a limited number of studies were found to address the three research questions raised in the introduction, especially the first one. This reflects the fact that research focusing specifically on disruptiveness in preschoolers is still sparse. We found several studies in which participants included preschoolers, school-age children, and adolescents in the same analyses. Those studies could not be included in our review, however, as their findings may or may not have been applicable to preschoolers. Although we are encouraged by the consistency of our findings, they are tentative given the small number of studies available for review. The literature would benefit from studies aimed at replicating and contributing to the results reported here and, when samples vary greatly in age, from detailed reports that break down findings for preschoolers, school-age children, and adolescents.

Fourth, diversity of the samples included in this review was limited, especially with respect to child gender and ethnicity. Although some studies failed to report demographic information, we estimate that the overall sample consisted primarily of boys (67%, with four studies including boys only) of mainstream American culture (68% Caucasian, 17% African American, and 15% Other, i.e., Hispanic, Asian, or Biracial). Future studies will need to recruit more diverse samples to determine the extent to which categorical and dimensional approaches adequately assess early disruptiveness in boys and girls, and yield comparable results for children of different ethnic backgrounds. Of the 26 studies included in this meta-

analysis, none broke down results by child gender or ethnicity, making it impossible to estimate the extent to which the overall findings are consistent across groups. Future studies will also need to report much more detailed demographic information about their samples. Of the 26 studies reviewed, 8 did not report any information on child ethnicity and the majority provided no data on parental age, marital status, income, or education.

Finally, several of the studies included in this review reported a very limited number of statistics needed to conduct a meta-analysis (often omitting basic statistics, such as t , F , or χ^2 values). For example, one study focusing on the stability of early disruptiveness only included the sample N at each time point and the percentage of children who continued to display disruptive behavior problems at follow-up (Campbell, 1994). Although statistical programs such as DSTAT allow for calculation of effect sizes using a variety of statistics, greater confidence can be placed in the results of studies that provide comprehensive statistical data.

As these limitations show, the literature on early disruptiveness still lacks the “ideal” study necessary to distinguish disruptive behavior problems from normative child behavior in the preschool years and to outline their developmental trajectory. Such a study would address the three research questions examined in the current review and include at least two groups of children (referred and non-referred for disruptive behavior), both assessed at baseline and follow-up using more than one categorical and one dimensional measures of disruptiveness. The study would include boys and girls from diverse ethnic and socioeconomic backgrounds, in sufficient numbers to provide adequate statistical power to conduct subgroup analyses. Finally, the study would report detailed demographic information about the sample (children and their parents/families) and include sufficient statistical data to be included in a future meta-analysis. Until comprehensive studies of this kind are conducted, this and similar reviews can only point to important gaps in knowledge as they summarize what is already known in this area.

4.2. Conclusion

Taken together, results of this meta-analysis provide evidence that, in the measurement of early disruptiveness, categorical and dimensional approaches: (1) yield comparable data; (2) demonstrate that early disruptiveness is stable over time; and (3) distinguish between disruptive preschoolers referred for services and their non-referred peers. In spite of important limitations in the literature, this suggests that disruptive behavior problems can be identified as early as age 3 and that both approaches may be used interchangeably and/or simultaneously for that purpose.

Acknowledgements

This study was supported by grant R21 HD40079 from the National Institute of Child Health and Human Development and by grant R49/CCR 522339 from the Centers for Disease Control and Prevention to the second author.

References*

- Achenbach T, & Edelbrock C (1983). *Manual for the Child Behavior Checklist and Revised Child Behavior Profile* Burlington: University of Vermont Department of Psychiatry.
- *Achenbach T, Edelbrock C, & Howell C (1987). Empirically based assessment of the behavioral/problems of 2- and 3-year old children. *Journal of Abnormal Child Psychology*, 15(4), 629–650. [PubMed: 3437096]
- *Alink L, Mesman J, Zeijl J, Stolk M, Juffer F, Koot H, et al. (2006). The early childhood aggression curve: Development of physical aggression in 10–50 month children. *Child Development*, 77, 954–966. [PubMed: 16942499]
- American Psychiatric Association (2000). *Diagnostic and Statistical Manual for Mental Disorders—IV TR*, 4th edition—Text revision. Washington, DC: American Psychiatric Association.
- Arend R, Lavigne J, Rosenbaum D, Binns H, & Christoffel K (1996). Relationship between taxonomic and quantitative diagnostic systems in preschool children: Emphasis on disruptive disorders. *Journal of Clinical Child Psychology*, 25(4), 388–397.
- Beauchaine T (2003). Taxometrics and developmental psychopathology. *Development and Psychopathology*, 15, 501–527. [PubMed: 14582930]
- Braswell L, August G, Bloomquist M, & Realmuto G (1997). School-based secondary prevention for children with disruptive behavior: Initial outcomes. *Journal of Abnormal Child Psychology*, 25(3), 197–208. [PubMed: 9212372]
- *Campbell S (1987). Parent-referred problem 3-year-olds: Developmental changes in symptoms. *Journal of Child Psychology and Psychiatry*, 28(6), 835–845. [PubMed: 3436992]
- *Campbell S (1990). *Behavior problems in preschool children* New York: Guilford.
- *Campbell S (1994). Hard-to-manage preschool boys: Externalizing behavior, social competence, and family context at 2-year follow-up. *Journal of Abnormal Child Psychology*, 22(2), 147–166. [PubMed: 8064027]
- *Campbell S, & Ewing L (1990). Follow-up of hard-to-manage preschoolers: Adjustment of age 9 and predictors of continuing symptoms. *Journal of Child Psychology and Psychiatry*, 31(6), 871–889. [PubMed: 2246339]
- *Campbell S, Ewing L, Breaux A, & Szumowski E (1986). Parent-referred problem three year-olds: Follow-up at school entry. *Journal of Child Psychology and Psychiatry*, 27(4), 473–488. [PubMed: 3745347]
- Cohen J (1983). The cost of dichotomization. *Applied Psychological Measurement*, 7, 249–253.
- Cohen J (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- *Crawford L, & Lee S (1991). Test–retest reliability of the Child Behavior Checklist ages 2–3. *Psychological Reports*, 69, 496–498.
- Dodge KA, Coie JD, & Lynam D (2006). Aggression and antisocial behavior in youth. In Damon W, Lerner RM, & Eisenberg N (Eds.), *Handbook of Child Psychology: Vol. 3. Social, Emotional, and Personality Development* (6th ed.) (pp. 719–788). New York: Wiley.
- Dumas JE (1996). Why was this child referred? Interactional correlates of referral status in families of children with disruptive behavior problems. *Journal of Clinical Child Psychology*, 25, 106–116.
- Dumas JE, Nissley-Tsiopinis J, & Moreland AD (2006). From intent to enrollment, attendance, and participation in preventive parenting groups. *Journal of Child and Family Studies*, 16, 1–26.
- Earls F (1980). The prevalence of behavior problems in 3-year-old children. *Archives of General Psychiatry*, 52, 1153–1159.
- *Egeland, Kalkoske M, Gottesman N, & Erickson M (1990). Preschool behavior problems: Stability and factors accounting for change. *Journal of Child Psychology and Psychiatry*, 31(6), 891–909. [PubMed: 2246340]
- Eyberg S, Schuhmann E, & Rey J (1998). Child and adolescent psychotherapy research: Developmental issues. *Journal of Abnormal Child Psychology*, 26, 71–82. [PubMed: 9566547]

- *Fagot B, & Leve L (1998). Teacher ratings of externalizing behavior at school entry for boys and girls: Similar early predictors and different correlates. *Journal of Child Psychology and Psychiatry*, 39(4), 555–566. [PubMed: 9599783]
- *Gadow K, Sprafkin J, & Nolan E (2001). DSM-IV symptoms in community and clinic preschool children. *Journal of the American Academy of Child and Adolescent Psychiatry*, 40(12), 1383–1392. [PubMed: 11765283]
- Gillion M, Shaw D, Beck J, Schonberg M, & Lukon J (2002). Anger regulation in disadvantaged preschool boys: Strategies, antecedents, and the development of self-control. *Developmental Psychology*, 38, 222–235. [PubMed: 11881758]
- Hedges L, & Olkin I (1985). *Statistical Methods for Meta-analysis* San Diego: Academic Press.
- *Heller T, Baker B, Henker B, & Hinshaw S (1996). Externalizing behavior and cognitive functioning from preschool to first grade: Stability and predictors. *Journal of Clinical Child Psychology*, 25(4), 376–387.
- Johnson BT (1989). *DSTAT: Software for the Meta-Analytic Review of Research Literatures* Hillsdale, NJ: Lawrence Erlbaum Associates.
- Kazdin AE (2005). *Parent management training* New York: Oxford University Press.
- Keenan K, Shaw D, Delliquadri E, Giovannelli J, & Walsh B (1998). Evidence for the continuity of early problem behaviors: Application of a developmental model. *Journal of Abnormal Child Psychology*, 26(6), 441–454. [PubMed: 9915651]
- *Keenan K, & Wakschlag L (2000). More than the terrible twos: The nature and severity of behavior problems in clinic-referred preschool children. *Journal of Abnormal Child Psychology*, 28, 33–46. [PubMed: 10772348]
- *Keenan K, Wakschlag L, Danis B, Hill C, Humphries M, & Duax J (2007). Further evidence of the reliability and validity of DSM-IV ODD and CD in preschool children. *Journal of the American Academy of Child and Adolescent Psychiatry*, 46, 433–435. [PubMed: 17420674]
- *Lahey B, Pelham W, Loney J, Lee S, & Willcutt E (2005). Instability of the DSM-IV subtypes of ADHD from preschool through elementary school. *Archives of General Psychiatry*, 62, 896–902. [PubMed: 16061767]
- Lavigne J, Arend R, Rosenbaum D, Binns H, Christoffel K, & Gibbons R (1998). Psychiatric disorders with onset in the preschool years: I. Stability of diagnosis. *Journal of the American Academy of Child and Adolescent Psychiatry*, 37, 1246–1254. [PubMed: 9847496]
- Lavigne J, Binns H, Arend R, Rosenbaum D, Christoffel K, Hayford J, et al. (1998). Psychopathology and healthcare use among preschool children: A retrospective analysis. *Journal of the American Academy of Child and Adolescent Psychiatry*, 37, 262–270. [PubMed: 9519630]
- Lavigne J, Gibbons R, Christoffel K, Arend R, Rosenbaum D, Binns H, et al. (1996). Prevalence rates and correlates of psychiatric disorders in children. *Journal of the American Academy of Child and Adolescent Psychiatry*, 32(2), 204–214.
- Loeber R (1982). The stability of antisocial and delinquent child behavior: A review. *Child Development*, 53, 1431–1446. [PubMed: 6756808]
- Loeber R (1990). Development and risk factors of juvenile antisocial behavior and delinquency. *Clinical Psychology Review*, 10, 1–41.
- *Marakovitz S, & Campbell S (1998). Inattention, impulsivity, and hyperactivity from preschool to school age: Performance of hard-to-manage boys on lab measures. *Journal of Child Psychology and Psychiatry*, 39(6), 841–851. [PubMed: 9758193]
- Masse L, & Tremblay R (1999). Kindergarten disruptive behavior, family adversity, gender, and elementary school failure. *International Journal of Behavioural Development*, 23(1), 225–240.
- Meehl P (2001). Comorbidity and taxometrics. *Clinical Psychology: Science and Practice*, 8, 507–519.
- *Mesman J, & Koot H (2001). Early preschool predictors of preadolescent internalizing and externalizing DSM-IV diagnoses. *Journal of the American Academy of Child and Adolescent Psychiatry*, 40, 1029–1036.
- *Mesman J, Bongers I, & Koot H (2001). Preschool developmental pathways to preadolescent internalizing and externalizing problems. *Journal of Child Psychology and Psychiatry*, 5, 679–689.

- *Olson S, & Brodfeld P (1991). Assessment of peer rejection and externalizing behavior problems in preschool boys: A short-term longitudinal study. *Journal of Abnormal Child Psychology*, 19(4), 493–503. [PubMed: 1757714]
- *Owens E, & Shaw D (2003). Predicting growth curves of externalizing behavior across the preschool years. *Journal of Abnormal Child Psychology*, 31, 575–590. [PubMed: 14658739]
- *Pelletier J, Collett B, Gimpel G, & Crowley S (2006). Assessment of disruptive behavior in preschoolers: Psychometric properties of the Disruptive Behavior Disorders Rating Scale and the School Situations Questionnaire. *Journal of Psychoeducational Assessment*, 24(1), 3–18.
- *Pierce E, Ewing L, & Campbell S (1999). Diagnostic status and symptomatic behavior of hard-to-manage preschool children in middle childhood and early adolescence. *Journal of Clinical Child Psychology*, 28, 44–57. [PubMed: 10070606]
- *Querido J, & Eyberg S (2003). Psychometric properties of the Sutter–Eyberg Student Behavior Inventory—Revised with preschool children. *Behavior Therapy*, 34, 1–15.
- Rosenthal R (1979). The “file drawer problem” and tolerance for null results. *Psychological Bulletin*, 86(3), 638–641.
- Scholte E, Van Berckelaer-Onnes I, & Van Der Ploeg J (2002). Factorial validity, reliability of assessments and prevalence of ADHD behavioural symptoms in day and residential treatment centres for children with behavioural problems. *International Journal of Methods in Psychiatric Research*, 11 (1), 33–44. [PubMed: 12459803]
- Shaw D, Owens E, Giovannelli J, & Winslow E (2001). Infant and toddler pathways leading to early externalizing disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, 40, 36–43. [PubMed: 11195559]
- *Speltz M, McClellan J, DeKlyen M, & Jones K (1999). Preschool boys with oppositional defiant disorders: Clinical presentation and diagnostic change. *Journal of the American Academy of Child and Adolescent Psychiatry*, 38(7), 838–845. [PubMed: 10405501]
- *Sprafkin J, Volpe R, Gadow K, Nolan E, & Kelly K (2002). A DSM-IV referenced screening instrument for preschool children: The Early Childhood Inventory-4. *Journal of the American Academy of Child and Adolescent Psychiatry*, 41(5), 604–612. [PubMed: 12014793]
- Teagarden L, & Burns G (1999). 12-month stability of ADHD, opposition defiant disorder, and conduct disorder symptoms in kindergarten through fifth grade children based on a single information source. *Child and Family Behavior Therapy*, 21(3), 53–70.
- Vondra J, Shaw D, Swearingen L, Cohen M, & Owens E (2001). Attachment stability and emotional and behavioral regulation from infancy to preschool age. *Development and Psychopathology*, 13, 13–33. [PubMed: 11346048]
- *Wakschlag L, & Keenan K (2001). Clinical significance and correlates of disruptive behavior in environmentally at-risk preschoolers. *Journal of Clinical Child Psychology*, 30(1), 262–275. [PubMed: 11393926]
- Waldman I, Lilienfeld S, & Lahey B (1995). Toward construct validity in the childhood disruptive behavior disorders. *Advances in Clinical Child Psychology*, 17, 323–363.
- White J, Moffitt T, Earls T, & Robbins L (1990). Preschool predictors of persistent conduct disorder and delinquency. *Criminology*, 28, 443–454.
- Wolf F (1986). *Meta-analysis: Quantitative Methods for Research Synthesis*. (Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 07–059) Newbury Park, CA: Sage Publications Inc.
- Wortman P (1994). Judging research quality In Cooper H & Hedges L (Eds.), *The Handbook of Research Synthesis* (pp. 97–110). New York: Russell Sage Foundation.
- Zahn-Waxler C, Ianotti R, Cummings E, & Denham S (1990). Antecedents of behavior problems in children of depressed mothers. *Developmental Psychopathology*, 2, 271–291.

Table 1

Descriptive statistics of reviewed articles

Article	N	Child gender	Child age	Child ethnicity ¹	Parent age	Marital status ²	Income level	Parent education	Clinical/ community	Setting
Achenbach et al. (1987)	87	–	–	–	–	–	–	–	–	–
Alink et al. (2006)	744	390 (52%) boys 354 (48%) girls	<i>M</i> = 2.98 (0.09) Range = 2.75– 3.33	“Mostly” EA	–	–	–	–	Community	–
Campbell et al. (1986)	68	41 (60%) boys 27 (40%) girls	<i>M</i> = 2.92 Range = 2.00– 3.00	–	–	–	“Range in social class”	–	Parent referred	–
Campbell and Ewing (1990)	54	36 (67%) boys 18 (33%) girls	–	–	–	46 (85%) M 8 (15%) S	–	–	Parent referred	Urban
Campbell (1987)	68	41 (60%) boys 27 (40%) girls	<i>M</i> = 2.92	–	–	–	–	–	Community	–
Campbell (1994)	112	112 (100%) boys	<i>M</i> = 3.83 Range = 2.42– 4.83	110 (98%) EA	–	–	“Working and middle class”	–	School	–
Crawford and Lee (1991)	30	14 (47%) boys 16 (53%) girls	Range = 2.00– 3.00	25 (83%) EA 2 (7%) H 3 (8%) Other	–	–	–	–	School	Rural
Egeland et al. (1990)	96	52 (54%) boys 44 (46%) girls	Range = 4.50– 5.00	–	–	–	<i>M</i> = \$14,500	–	School	–
Fagot and Leve (1998)	156	82 (53%) boys 74 (47%) girls	–	148 (95%) EA 3 (2%) AA 2 (1%) H 3 (2%) Other	–	137 (88%) M	<i>M</i> = \$15,000	–	Community	–
Gadow et al. (2001)	755	443 (59%) boys 312 (41%) girls	<i>M</i> = 4.25 (0.75)	504 (67%) EA 70 (9%) AA 49 (6%) H 7 (1%) Other	–	–	–	–	Outpatient clinic	Urban
Heller et al. (1996)	77	37 (48%) boys 40 (52%) girls	<i>M</i> = 4.60 (0.77)	53 (69%) EA	<i>M</i> = 37	12 (16%) S	67 (87%) Middle class or above	–	Community	–
Keenan and Wakschlag (2000)	79	61 (77%) boys 18 (23%) girls	<i>M</i> = 4.00(0.75) Range = 3.00– 4.11	Range = 24– 49 10 (13%) EA 63 (80%) AA 1 (1%) H	–	–	Low income (82% on welfare)	–	Inpatient clinic	Urban

Article	N	Child gender	Child age	Child ethnicity ¹	Parent age	Marital status ²	Income level	Parent education	Clinical/ community	Setting
Keenan et al. (in press)	100	56 (56%) boys 44 (44%) girls	Range= 3.00– 5.00	5 (5%) Other 82 (82%) AA 6 (6%) H	–	–	–	–	Outpatient clinic	Urban
Lahey et al. (2005)	125	98 (78%) boys 20 (22%) girls	Range= 3.80– 7.00	74 (56%) EA 36 (29%) AA 8 (6%) Other	–	–	–	–	Outpatient clinic	Urban
Mesman, Bongers, and Koot (2001)	397	204 (51%) boys 193 (49%) girls	<i>M</i> = 5.31 (0.64)	–	–	345 (87%) M 52 (13%) S	–	–	Community	Urban
Mesman and Koot (2001)	420	215 (51%) boys 205 (49%) girls	<i>M</i> = 2.60 (0.80)	–	–	–	–	–	Community	–
Marakovitz and Campbell (1998)	112	112 (100%) boys	<i>M</i> = 3.83	–	–	–	–	–	School	–
Olson and Brodfield (1991)	53	53 (100%) boys	<i>M</i> = 4.60 Range= 4.00– 5.50	51 (96%) EA	–	–	Low income	–	School	–
Owens and Shaw (2003)	275	275 (100%) boys	–	160 (58%) EA 110 (40%) AA 5 (2%) Other	–	179 (65%) M 184 (67%) poverty level	109 (76%) Unemployed;	160 (58%) no HS	–	Urban
Pelletier et al. (2006)	200	99 (50%) boys 101 (50%) girls	<i>M</i> = 4.60 (0.50)	–	–	–	–	–	School	–
Pierce et al. (1999)	59	–	<i>M</i> = 2.92	–	–	–	–	–	Community	–
Querido and Eyberg (2003)	74	40 (54%) boys 34 (46%) girls	<i>M</i> = 5.01 (1.37) Range= 3.20– 6.10	55 (74%) EA 9 (12%) AA 6 (9%) H 4 (5%) Other	–	62 (84%) M	Range= \$21,000– \$30,000	<i>M</i> = 14.69 (2.02)	School	Rural
Speltz, McClellan, DeKlyen, and Jones (1999)	92	92 (100%) boys	<i>M</i> = 4.73 (0.52)	76 (79%) EA 3 (7%) AA 13 (14%) Other	–	59 (64%) M	–	–	Outpatient clinic	Urban
Sprafkin et al. (2002)	224	172 (77%) boys 52 (23%) girls	<i>M</i> = 4.55 (0.77) Range= 3.00– 6.00	195 (87%) EA 13 (6%) AA 13 (6%) H 2 (1%) Other	–	–	–	–	Outpatient clinic	Rural
Wakschlag and Keenan (2001)	79	59 (75%) boys	<i>M</i> = 4.00 (0.78)	6 (8%) EA	–	–	59 (88%) on welfare	–	Inpatient clinic	Urban

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Article	N	Child gender	Child age	Child ethnicity ¹	Parent age	Marital status ²	Income level	Parent education	Clinical/ community	Setting
	20 (25%) girls	Range= 3.00– 4.11	63 (80%) AA 6 (8%) H 3 (4%) Other							

¹EA = European American, AA = African American, H = Hispanic, Other = Asian, Native American, or Biracial;

²M = married, S = single, -- = Information not reported.

Table 2

Individual effect sizes

Article	Measure	<i>r</i>	Effect size (CI)	Size	<i>r</i>	Effect size (CI)	Size	Effect size (CI)	Size
chenbach, Edelbrock, and Howell (2004)	Child Behavior Checklist	DIM			0.70	1.96 (1.60/2.32)	L	0.40	0.87 (0.57/1.18) L
Alink et al. (2006)–Study 1	Physical Aggression Scale for Early Child	DIM			0.63	1.62 (1.27/1.97)	L		
Alink et al. (2006)–Study 2	Physical Aggression Scale for Early Child	DIM			0.72	2.08 (1.71/2.44)	L		
Campbell (1987)	Child Behavior Checklist	DIM			0.66	1.77 (1.32/2.22)	L		
Campbell (1994)	Swanson, Pelham, and Nolan Questionnaire	C-AD			0.53	1.25 (0.54/1.97)	L	0.51	1.19 (0.65/1.73) L
Campbell (1994)	Swanson, Pelham, and Nolan Questionnaire	C-OD						0.41	0.91 (0.38/1.43) L
Campbell and Ewing (1990)	Child Behavior Checklist	DIM			0.46	1.04 (0.61/1.46)	L	0.49	1.13 (0.44/1.82) L
Campbell and Ewing (1990)	Child Behavior Checklist	C-AD			0.86	2.69 (2.14/3.25)	L	0.59	1.46 (0.73/2.18) L
Campbell et al. (1986)	Child Behavior Checklist	DIM			0.58	1.41 (0.87/1.95)	L		
Crawford and Lee (1991)	Child Behavior Checklist	DIM			0.74	2.00 (1.56/2.84)	L		
Egeland et al. (1990)	Child Behavior Checklist	DIM			0.33	0.70 (0.13/1.27)	M		
Fagot and Leve (1998)	Child Behavior Checklist	DIM			0.41	0.90 (0.67/1.13)	L		
Gadow et al. (2001)	Early Childhood Inventory-4	C-AD			0.46	1.04 (0.79/1.29)	L	0.51	1.20 (0.99/1.41) L
Gadow et al. (2001)	Early Childhood Inventory-4	C-OD			0.56	1.35 (1.09/1.61)	L	0.39	0.86 (0.66/1.06) L
Heller et al. (1996)	Child Behavior Checklist	DIM			0.60	1.50 (1.14/1.86)	L	0.53	1.24 (0.51/1.95) L
Keenan and Wakschlag (2000)	K-SADS-Epidemiological 5th ver.	C-OD						0.29	0.61 (0.24/0.97) M
Keenan et al. (in press)	Kiddie-Disruptive Behavior Disorder Schedule	C-AD						0.83	2.97 (2.59/3.35) L
Lahey et al. (2005)	Disruptive Behavior Disorders Rating Scale	C-AD			0.23	0.47 (0.21/0.73)	M		
Marakovitz and Campbell (1998)	Swanson, Pelham, and Nolan Questionnaire	C-AD						0.26	0.54 (0.27/0.81) M
Mesman et al. (2001)	Child Behavior Checklist	DIM			0.01	0.02 (−0.19/0.23)	S		
Mesman and Koot (2001)	Child Behavior Checklist	DIM			0.09	0.17 (0.02/0.32)	S		
Olson and Brodfeld (1991)	Conners Teacher Questionnaire	C-OD			0.80	2.67 (2.14/3.19)	L		
Owens and Shaw (2003)	Child Behavior Checklist	DIM			0.39	0.86 (0.68/1.03)	L		
Pelletier et al. (2006)	Disruptive Behavior Disorders Rating Scale	AD	0.77	2.41 (2.16/ 2.67) L					
Pelletier et al. (2006)	Disruptive Behavior Disorders Rating Scale	OD	0.69	1.91 (1.67/ 2.14) L					
Pierce et al. (1999)	Swanson, Pelham, and Nolan Questionnaire	C-AD			0.32	0.67 (0.07/1.28)	M		
Pierce et al. (1999)	Swanson, Pelham, and Nolan Questionnaire	DB						0.34	0.72 (0.30/1.14) M
Pierce et al. (1999)	Swanson, Pelham, and Nolan Questionnaire	DB						0.77	2.41 (1.54/3.28) L

Article	Measure		<i>r</i>	Effect size (CI)	Size	<i>r</i>	Effect size (CI)	Size	<i>r</i>	Effect size (CI)	Size
Querio and Eyberg (2003)	Disruptive Behavior Disorders Rating Scale	DB							0.45	0.99 (0.61/1.38)	L
Speltz et al. (1999)	Child Behavior Checklist	DIM				0.18	0.37 (0.08/0.67)	S			
Sprafkin et al. (2002)	Early Childhood Inventory-4	AD	0.70	1.96 (1.71/2.21)	L				0.13	0.27 (0.02/0.56)	
Sprafkin et al. (2002)	Early Childhood Inventory-4	OD	0.81	2.76 (2.48/3.05)	L				0.20	0.42 (0.13/0.70)	S
Wakschlag and Keenan (2001)	K-SADS-Epidemiological 5th ver.	C-OD							0.74	2.20 (1.75/2.64)	L

IM = Dimensional Scale; C-AD = Categorical Scale—ADHD; C-ODD = Categorical Scale—ODD; EXT = externalizing; AD = ADHD; OD = ODD; L = large effect; M = medium effect; S = small effect