2006

Classroom Observation Of Children With Adhd And Their Peers: A Meta-analytic Review

Michael Kofler
University of Central Florida

Part of the Psychology Commons
Find similar works at: https://stars.library.ucf.edu/etd
University of Central Florida Libraries http://library.ucf.edu

This Masters Thesis (Open Access) is brought to you for free and open access by STARS. It has been accepted for inclusion in Electronic Theses and Dissertations, 2004-2019 by an authorized administrator of STARS. For more information, please contact STARS@ucf.edu.

STARS Citation
https://stars.library.ucf.edu/etd/4464
CLASSROOM OBSERVATION OF CHILDREN WITH ADHD AND THEIR PEERS:
A META-ANALYTIC REVIEW

by

MICHAEL J. KOFLER
B.S. Tulane University, 2000

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in the Department of Psychology in the College of Arts and Sciences at the University of Central Florida Orlando, Florida

Spring Term
2006
ABSTRACT

Meta-analysis of 23 between-group direct observation studies of children with ADHD and typically developing peers indicates significant deficiencies in children with ADHD's ability to pay attention in classroom settings. Comparison with 59 single case design studies of children with ADHD suggests generalizability of between-group comparisons. Weighted regression analysis determined that several methodological differences – sample characteristics, diagnostic procedures, and observational coding schema – have significant effects on observed levels of attentive behavior in the classroom. Best case estimation indicates that after accounting for these factors, children with ADHD are on-task approximately 65% of the time compared to 85% for their classroom peers. Children with ADHD were also more variable in their attentive behavior across studies. Implications for conceptual models of ADHD are discussed.
# TABLE OF CONTENTS

LIST OF FIGURES ....................................................................................................................... vi

LIST OF TABLES ........................................................................................................................ vii

LIST OF ACRONYMS/ABBREVIATIONS .............................................................................. viii

CHAPTER ONE: INTRODUCTION ............................................................................................. 1
  The Need for a Meta-Analytic Approach ................................................................................... 4
  Sample Characteristics ............................................................................................................ 6
  Diagnostics Methods ............................................................................................................... 6
  Classroom variables .............................................................................................................. 7
  Observation methods ............................................................................................................. 8

CHAPTER TWO: METHODOLOGY .......................................................................................... 10
  Literature Searches ................................................................................................................ 10
  Inclusion and Exclusion Criteria ............................................................................................ 11
  Coding of Moderators ........................................................................................................... 12
    Sample ............................................................................................................................... 12
    Diagnostics ........................................................................................................................ 12
    Classroom variables ......................................................................................................... 14
    Observation variables ....................................................................................................... 14
  Computation of Effect Sizes ................................................................................................ 16
    Hedges' g ............................................................................................................................ 16
    Multiple effect sizes ......................................................................................................... 17
    Nonsignificant results ....................................................................................................... 17
Data Analysis ........................................................................................................................................... 18

CHAPTER THREE: RESULTS ................................................................................................................ 21

Tier I: Moderator-independent attentional differences ........................................................................... 21

Publication Bias: The File Drawer Problem ...................................................................................... 21

Off-task Comparisons ..................................................................................................................... 21

Effect sizes .......................................................................................................................................... 22

Tier II: Moderators of obtained effect size .......................................................................................... 22

Fixed Effects Model .......................................................................................................................... 22

Sample ............................................................................................................................................... 23

Diagnostics ......................................................................................................................................... 24

Classroom type .................................................................................................................................... 25

Observation Variables ...................................................................................................................... 25

Tier III: Best Case Estimation and Original Metric ............................................................................. 27

Best Case Estimation .......................................................................................................................... 27

Power Analysis for Future Studies .................................................................................................... 27

Original Metric .................................................................................................................................. 28

Tier IV: Single Case Design Studies .................................................................................................. 28

Mean Differences ................................................................................................................................. 28

Effect Size Comparisons ................................................................................................................... 29

Moderator Comparison ...................................................................................................................... 29

CHAPTER FOUR: CONCLUSION .......................................................................................................... 45

LIST OF REFERENCES ...................................................................................................................... 57
LIST OF FIGURES

Figure 1: Original Metric Differences Between Observed Classroom Attention Of Children With Adhd And Control
Children Based On Best Case Estimation After Controlling For Methodological Differences Among Studies.
-----------------------------------------------------------------------------------------------------------------------------.43
LIST OF TABLES

Table 1: Demographic And Methodological Variables .................................................................30

Table 2: Mean Off-Task Rates, Standard Difference Scores, And Effect Sizes In Children With Adhd ...........33

Table 3: Stem And Leaf Display Of 23 Hedges G Effect Sizes ............................................................35

Table 4: Inverse Variance Weighted Regression Results .................................................................36

Table 5: Predicted And Obtained Effect Sizes For Single Case Design Studies .........................................37

Table 6: Between Group And Single Case Design Study Moderator Variable Differences ..........................38

Table 7: Single Case Design Studies: Demographic And Methodological Variables ............................39
# LIST OF ACRONYMS/ABBREVIATIONS

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADHD</td>
<td>Attention-Deficit/Hyperactivity Disorder</td>
</tr>
<tr>
<td>BG</td>
<td>Between Group</td>
</tr>
<tr>
<td>CD</td>
<td>Conduct Disorder</td>
</tr>
<tr>
<td>CPT</td>
<td>Continuous Performance Task</td>
</tr>
<tr>
<td>DSM-IV</td>
<td>Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition</td>
</tr>
<tr>
<td>ES</td>
<td>Effect Size</td>
</tr>
<tr>
<td>LD</td>
<td>Learning Disabilities</td>
</tr>
<tr>
<td>ODD</td>
<td>Oppositional Defiant Disorder</td>
</tr>
<tr>
<td>SCD</td>
<td>Single Case Design</td>
</tr>
</tbody>
</table>
CHAPTER ONE: INTRODUCTION

Attention-Deficit/Hyperactivity Disorder (ADHD) is a chronic and disabling condition that affects an estimated three to seven percent of school-age children (APA, 2000; Francis, 1993). Inattention in the classroom is associated with several negative outcomes: These children are academically more likely to receive lower grades and standardized test scores (Zentall, 1993), be placed in special education classrooms, have comorbid learning disabilities, and repeat a grade level compared to their typically developing peers (Faraone et al., 1993). More than half of all children with ADHD fail at least one grade level by adolescence (Barkley et al., 1990), and 23 percent of children with ADHD fail to finish high school (Mannuzza et al., 1993). Early academic difficulties are predictive of later academic problems (Fergusson & Horwood, 1995; Rapport, Scanlan, & Denney, 1999), and longitudinal outcome studies of ADHD indicate that adults diagnosed with ADHD as children are less likely to attend college, and more likely to have unstable marriages, lower mean SES scores, and conduct problems compared to peers of similar intelligence (Mannuzza et al., 1993).

Environmental demands impact the expression of ADHD-related symptoms (Rapport, Chung, Shore, Denney, & Isaacs, 2000). Children with ADHD are indistinguishable from their peers in some environments (e.g., during recess and non-academically related work periods), but exhibit marked behavioral differences when encountering particular structured situational demands (Porrino et al., 1983; Whalen et al., 1978). They pay attention less often during academic instruction (Abikoff, Gittelman-Klein, & Klein, 1977), complete academic work inconsistently (Barkley, DuPaul, & McMurray, 1990), and shift from one activity to another at home more frequently compared with peers (DuPaul & Stoner, 2004). Situations involving novel
tasks (Power, 1992; Zentall & Meyer, 1987) and/or frequent reinforcement (Douglas & Parry, 1983; Pelham, Milich, & Walker, 1986) are associated with improved behavior and performance in children with ADHD.

Laboratory studies employing rigorous experimental paradigms such as the Continuous Performance Task (CPT) often, yet not always, report performance differences between children with ADHD and typically developing children (Chung, Rapport, & Denney, 2005). The highly controlled nature of the laboratory setting allows hypotheses to be tested regarding the underlying processes of ADHD. However, such situational demands may evoke attentional processes that differ from those required in the natural environment (Barkley, 1991).

Ecological validity concerns necessitate the investigation of more naturalistic procedures for characterizing behavior, such as behavioral rating scales and direct observations by independent researchers. Ratings scales are cost- and time-efficient measures capable of capturing overall behavioral functioning. They provide a quantitative metric of the frequency and severity of behavior, yet remain dependent upon subjective judgments and are vulnerable to response bias, misinterpretation of questions, and over/underestimation of behavior due to intensity/immediacy effects (Kazdin, 1997; Rutter & Graham, 1968). In contrast, direct observations can provide more objective information by utilizing independent, well-trained observers and specific, operationalized behavioral definitions (Kazdin, 1997). Observational coding schemes offer decreased threats to validity in comparison with other methods. Unlike global rating scales of behavior, observation codes are not subject to halo effects or rater expectation bias (Abikoff et al., 2002; Harris & Lahey, 1982; Kent, O'Leary, Diament, & Dietz, 1974). Systematic and clearly defined behavior codes also may offer improved criterion validity (Abikoff et al., 1977). Observation coding schemes are able to differentiate between typically
developing children and children with ADHD, suggesting diagnostic validity (Luk, 1985; Pelham, Fabiano, & Massetti, 2005; Platzman et al., 1992). Classroom observation methods are more likely than laboratory methods to report significant attentional differences between children with ADHD and their peers (Platzman et al., 1992). However, sensitivity and specificity have not been sufficiently established to allow diagnosis based solely on an observation code, or any other single indicator (Abikoff et al., 2002; APA, 2000; Luk, 1985). Shortcomings of direct observation methods include increased time and labor commitments as well as a lack of agreement on behavioral definitions and standardized observation techniques.

Attentional problems are a core feature of ADHD, defined behaviorally in the DSM-IV by characteristics such as distractibility, organizational problems, daydreaming, and frequent shifting from one unfinished task to another (APA, 2000). Definitions of attention in the literature refer to preferential selection and processing of sensory information – inattention is inferred by frequent shifts in activity, and behaviors that are not task-related (Bear, Connors, & Paradiso, 2001; Platman et al., 1992). In direct observation studies, attention is most often operationalized as on-task behavior (e.g., Abikoff et al., 2002; DuPaul & Rapport, 1993). Definitions invariably involve the child physically looking at the teacher or seat work, but vary considerably in the duration of consecutive focus needed to code the child as on- or off-task. This operationalization is limited by the need for observable behavior, and it is noted that it is possible to pay attention to auditory stimuli without visually focusing, just as the direction of one's eyes may not indicate task attention (e.g., daydreaming). The former may be more relevant in studies observing children during teacher-led lectures, whereas the latter may influence results of studies observing children during structured seat work. In laboratory settings, on-task data has also been used to characterize vigilance (e.g., omission errors in CPT studies; Chung et al.,
In a natural classroom setting, however, vigilance is only one of myriad factors influencing a child's ability to pay attention or remain on-task (DuPaul & Stoner, 2004; Whalen et al., 1978). In addition, behavioral observations of children with ADHD performing a CPT task are more highly correlated with direct observations of classroom on-task behavior than CPT performance (Schatz, Ballantyne, & Trauner, 2001).

The Need for a Meta-Analytic Approach

Attentional problems in the classroom are quintessential symptoms of children with ADHD, and are often the catalyst for clinical referrals (Pelham, Fabiano, & Massetti, 2005). A meta-analytic review provides not only documentation of this phenomenon across studies, but unequivocal quantification of the magnitude of these attentional difficulties compared to typically developing children. It is an accepted tenet of attention-deficit/hyperactivity disorder (ADHD) that affected children are off-task more frequently and/or for longer durations than their peers. However, the magnitude of this difference varies considerably across studies. Two previous review articles have found that the overwhelming majority of direct observation studies report significant differences between children with ADHD and their peers (Luk, 1985; Platzman et al., 1992). Both reviews described classroom and laboratory observation studies and found that most studies reported significant differences between the attentive behaviors of ADHD and normal children. Luk (1985) concluded that differences in task demands and classroom situations influenced behavioral differences between hyperactive and comparison children, but noted that both the specific factors and their relative effects on behavior remain unknown. Platzman and colleagues (1992) used difference-of-proportion tests to examine whether levels of a particular variable are more or less likely to result in significant between-group differences. They
concluded that significant between-group differences were more likely to occur using classroom rather than laboratory observation, however, diagnostic source (i.e., parent, teacher, or physician report) was unrelated to the number of studies reporting significant differences. Both reviews noted the significant variability among behavioral coding schemes, participant age, and study setting (Luk, 1985; Platzman et al., 1992). The reviews, however, were primarily descriptive in nature and failed to quantify between-study differences or analyze potential moderator variable effects on observed differences in attention between children with ADHD and typically developing controls. Box score literature reviews (e.g., comparing the number of studies finding versus failing to find significant group differences) do not consider study power, and results may therefore inaccurately reflect the data (see Howard, Maxwell, & Fleming, 2000 for details and specific examples of this phenomenon). Moderating variables warrant scrutiny because of their potential to change the nature of dependent-independent variable relationships, with implications for theory development, refinement, and refutation (Holmbeck, 1997).

The present meta-analysis of published and unpublished studies examines both the magnitude and variability of observed differences in classroom attention (on-task behavior) between children with ADHD and their peers using standardized effect size estimates. The potential moderating relationship of sample characteristics, diagnostic methods, classroom variables, and observational schema on the magnitude of observed attentional differences in the classroom are discussed below and analyzed to determine whether the variability between effect size estimates exceeds levels expected based on study-level sampling error.
Sample Characteristics

Despite the restricted range related to the limited number of female subjects in most studies, gender differences in ADHD prevalence rates and theoretical differences in symptom manifestation necessitated the inclusion of this variable in moderator analysis (Barkley, 1990). Gender has been found to predict DSM-IV diagnostic subtype in most but not all studies of children with ADHD, with females more likely to exhibit inattentive symptoms and males more likely to display hyperactive behavior (Abikoff et al., 2002; Biederman & Faraone, 2004; Graetz, Sawyer, & Baghurst, 2005; Newcorn et al., 2001; Seidman, Biederman, & Monuteaux, 2005; Weiss, Worling, & Wasdell, 2003; Yang, Jong, & Chung, 2004). Cognitive differences have also been found between girls and boys diagnosed with ADHD (e.g., Carlson, Lahey, & Neeper, 1986; Douglas, 1988). A recent large-scale study of nonreferred boys and girls, however, found no gender differences on any variables of interest. These results suggest that some earlier findings may be the result of referral bias rather than true gender differences in the expression of ADHD symptoms (Biederman et al., 2005). Previous meta-analytic reviews of children with physical disabilities, autism, and dyslexia have reported that matching on gender, age, and other demographic variables significantly influences outcome measures (Lavigne & Faier-Routman, 1992; Mottron, 2004; van Ijzendoorn & Bus, 1994).

Diagnostic Methods

Comprehensive clinical interviews, including structured/semi-structured interviews and data from multiple informants, are the gold standard for ADHD diagnosis, whereas rating scales provide a time- and cost-efficient method of identifying children whose behaviors are similar to those of children meeting formal diagnostic criteria (McClellan & Werry, 2000; Pelham, Fabiano, & Massetti, 2005; Rutter & Graham, 1968). Defining samples based solely on a referral
appears to be the least face valid method of grouping, considering the myriad of disorders and conditions featuring attention and behavioral problems as core or secondary features (APA, 2000). Extant research suggests that significant correlations exist between rating scale cutoff scores and semi-structured clinical interview diagnoses, as well as between parent and teacher rating scale scores (Biederman, Keenan, & Faraone, 1990; Hodges, 1993; McGrath, Handwerk, Armstrong, Lucas, & Friman, 2004). However, significant unexplained variability exists between both diagnostic tools and informants. For example, McGrath and colleagues (2004) reported correlations of .72 and .55 between symptom endorsement on the Diagnostic Interview Schedule for Children – Fourth Edition (DISC-IV; Schaffer, Fisher, Lucas, Dulcan, & Schwab-Stone, 2000) and Conners parent and teacher rating scales (Conners, Parker, Sitarenios, & Epstein, 1998), respectively. This finding suggests that between 48% and 70% of the variability in symptom endorsement on the semi-structured clinical interview is not accounted for by Conners rating scale data. The correlation between parent and teacher ratings is often reported to be somewhat lower than these values, highlighting the importance of considering informant source in the diagnostic process (e.g., McGrath et al., 2004).

Classroom variables

Situational variables have been found to affect hyperactive symptomatology in classroom (Barkley, Copeland, & Sivage, 1980; Beck, Kotkin, & Swanson, 1999; Pelham, Wheeler, & Chronis, 1998; Rapport, Murphy, & Bailey, 1982; Zentall, 1980) and other settings (e.g., Luk, 1985; Sleator & Ullmann, 1981). In addition, the specific academic task may have significant effects on the activity level and task attention of children with ADHD (e.g., Whalen et al., 1978; Zentall & Meyer, 1987).
Observation methods

Coding. The complexity of the coding scheme represents a trade-off between the amount of data collected and the potential for information processing limitations that may decrease data accuracy (Markman & Notarius, 1987). Coding fewer behaviors during live sessions is associated with improved reliability and interobserver agreement, and fewer recording errors (Dorsey, Nelson, & Hayes, 1986; Harris & Lahey, 1982). Smith, Madsen, and Cipani (1981) found no differences between continuous recording, interval recording, and recording by incident on measured reliability or observed rate of behavior. The interval type used by Smith and colleagues, however, did not correspond to any used by studies in the current analysis, whereas the continuous and by incident types descriptively matched. Other studies suggest improved reliability with the momentary (by incidence) compared to whole or partial intervals (Ary & Suen, 1983; Bramlett & Barnett, 1993), and others support the use of predefined intervals (Mehm & Knutson, 1987).

Duration. The influence of observation duration in the available literature is mixed. Observation duration is significantly related to variables of interest in some (Leaper, Anderson, & Sanders, 1998; Sahni, Schulze, & Stefanski, 1995) but not all (Gertz, Stilson, & Gynther, 1959; Rowley, 1978) studies in other areas. A previous meta-analytic review of expressive behavior failed to find outcome differences based on duration of observation (Ambady & Rosenthal, 1992), however, to our knowledge no studies of children with ADHD have directly examined the influence of observation duration on recorded behavior. Studies examining the time needed for reliable observational data, as measured by criterion codes and session intercorrelations, differ in their conclusions regarding the total number of minutes and days necessary. Some conclude generally that more observations are better, without specifying
minimum within- and between-day durations (Barton & Ascione, 1984; Leaper et al., 1998); others specify a minimum of five day, 20-minute sessions (Doll & Elliott, 1994) or 30-minute observations for three days (McKevitt & Elliott, 2005); and some fail to find an effect of observation duration on outcomes (Rowley, 1978). Both the population of interest and the frequency of target behavior likely influence the total observation duration needed for a reliable and valid sampling of behavior (Haynes, 2001). The behavior of children with ADHD has been described as consistently inconsistent, suggesting that longer observation periods may better capture the attentional processes of these children (e.g., Rapport et al., 1982).

Observational variables, including the type and quantity of behaviors measured (Lorber, 2004; Mezulis, Abramson, Hyde, & Hankin, 2004); and subject sampling, including diagnostic assessment method and informant, age, and gender, significantly moderate effect size estimates in most (Bogg & Roberts, 2004; Connell & Goodman, 2002; Sergerstrom & Miller, 2004) but not all (Lavigne & Faier-Routman, 1992) previous meta-analytic reviews in other areas.

Previous empirical studies support the influence of methodological variables, however, the relative impact of these variables for understanding classroom attentional differences in children with ADHD and typically developing children remains unknown (Luk, 1985). Significant moderating variables of these children’s classroom attention will be used to predict the overall mean effect size that would be expected if all direct observation studies employed the most rigorous methodology (Lipsey & Wilson, 2001). These metrics facilitate conclusions concerning overall differences in classroom attention between children with ADHD and their peers under ideal observation conditions.
CHAPTER TWO: METHODOLOGY

Literature Searches

A three-tier literature search was conducted using PsycInfo, PsycArticles, ERIC, Dissertation Abstracts International, and Social Science Citation Index. Search terms included permutations of the ADHD diagnostic label (ADHD, ADD, attention deficit, hyperactivity, hyperkinesis, minimal brain dysfunction/damage, MBD), class*, observ*, behav*, school, direct, attention, and on/off-task, where asterisks serve as wildcards (e.g., observ* will return studies with the word observation, observations, observer, etc.). Separate searches were conducted to find rating scale validation studies (additional search terms: rat*, reliab*, valid*, scale develop*) and single case design studies (additional search terms: functional analysis, case study, single subject). Searches were conducted with and without an ADHD search term included. Searches were conducted independently by two researchers (MJK and RMA), and repeated until no new studies were located. To further expand the initial study base, table of contents searches of the following journals likely to publish classroom observation studies of children with ADHD were undertaken: Journal of Applied Behavior Analysis, Behavior Modification, Behavior Research and Therapy, Behavior Therapy, Journal of School Psychology, School Psychology Review, and Journal of Attention Disorders. After the initial searches, studies cited by articles meeting inclusion criteria were examined (Tier II backward search), and a forward search (Tier III) was conducted using the Social Science Citation Index to locate studies citing those meeting inclusion criteria. These procedures generated 509 dissertations, peer-reviewed studies, and unpublished manuscripts written since 1962.
**Inclusion and Exclusion Criteria**

Inclusion/exclusion criteria were applied after the computerized searches were completed. They were not entered as search delimiters to avoid missing studies due to database misclassification. Inclusion and exclusion criteria are described below, with the number of studies omitted for each criterion in parentheses. The following served as inclusion criteria for the review: (a) an independent direct observation of children exhibiting inattentive, hyperactive, and/or impulsive behavior in an elementary classroom setting (204); (b) between six and 12 years of age (24); (c) on- or off-task frequency or duration data reported, or statistics reported in between group studies from which effect size can be estimated (94); and (d) low average or higher estimated intelligence (16). Exclusion criteria included: (a) comorbidity with other mental health disorders beyond Learning Disabilities (LD), Oppositional Defiant Disorder (ODD), or Conduct Disorder (CD) (11); (b) stimulant or psychotropic medication taken during observation, or no pretreatment baseline condition (74); and (c) repeat data (e.g., study published in journal and as book chapter; follow-up longitudinal study (4). Single case design studies were included if any subject(s) for which individual data was reported met inclusion criteria. Studies reporting only placebo (i.e., no medication-free baseline) conditions were excluded based on research demonstrating significant differences in functioning between baseline and placebo conditions in children with ADHD (e.g., Rapport, Denney, DuPaul, & Gardner, 1994). Twenty-three studies published from 1969 to 2004 met these search criteria and included a typically developing comparison group\(^1\). An additional 59 single case design studies published between 1962 and 2005 met the above criteria.

\(^1\) Special thanks to Carmen Himmerich for translating Lauth & Mackowiak (2004) from German.
Coding of Moderators

Demographic and methodological variables are shown in Table 1 and described below. Categorical variables were coded chronologically in the order listed below, such that higher values are associated with an addition to the variable in question (e.g., adding matched controls, diagnostic tools, observation time).

Sample

Total N. Sample size was used when both adjusting inflated effect sizes (i.e., Hedges' g correction) and weighting effect size contribution to analyses (i.e., inverse variance weights). Sample size was not coded as a moderator because further analysis would violate the variable independence assumption (Lipsey & Wilson, 2001).

Percent male. The proportion of male subjects was coded as (a) predominately or all male, or (b) 10% or more females, based on a previous meta-analytic review of predominately male samples (Mezulis et al., 2004).

Matching. Three matching variables were coded. Gender, age/grade, and classroom were each divided into two groups based on whether or not the researchers matched on each variable.

Diagnostics

Method. Diagnostic method was coded as: (a) referral for behavior problems with no further diagnostics; (b) single informant rating scale (i.e., parent or teacher); (c) multiple informant rating scales (i.e., parent and teacher); or (d) structured or semi-structured clinical interview. All but one study using a diagnostic structured/semi-structured clinical interview also included rating scales from multiple informants – the exception included a parent, but not a teacher rating scale (Roberts, 1990). All but four studies used at least one Conners (Conners et
al., 1998) or Achenbach (i.e., Child Behavior checklist/Teacher Report Form; Achenbach & Rescorla, 2001) rating scale. The four exceptions were all published before 1978.

Moniker. Diagnostic label was coded as (a) behavioral problem children; (b) hyperactive; or (c) ADD-H/ADHD. ADD-H and ADHD groups were combined due to research suggesting that children diagnosed as ADD-H based on the DSM-III typically meet DSM-IV ADHD Combined Type criteria (APA, 1987/2000; August & Garfinkel, 1993; Garfinkel & Amrami, 1992). Skansgaard and Burns (1998) reported separate results for DSM-IV Combined and Inattentive Types. The Combined Type data were used to match the subtype used in other included studies. The sole study not to specifically exclude children only meeting inattentive criteria was Atkins, Pelham, and Licht (1985), whose nonhyperactive subjects represent less than 20% of his total sample\(^2\). As expected, diagnostic moniker was significantly correlated with year of publication (\(r = .83\)). Inspection of the data reveals that only two studies (Book & Skeen, 1987; Roberts, 1990) fail to adhere to the following pattern: Studies published between 1969 and 1975 referred to their experimental group as Behavior Problem Children, research published between 1977 and 1980 referred to them as Hyperactive, journal articles from 1984 to 1993 used the ADD-H label, and studies from 1997 to 2004 use the current ADHD moniker. A correlation between year of publication and DSM version was not computed because only nine of the 23 studies explicitly reported the DSM version used. Publication year was not included in moderator analysis due to redundancy with diagnostic moniker (see footnote in results section).

\(^2\) Diagnostic subtypes are used in the DSM-III and DSM-IV, but not the DSM-III-R
Classroom variables

Type. Studies were coded into two categories: (a) normal classroom environments with the child's regular teacher and classmates; and (b) simulated classrooms in research settings.

Activities. The classroom activity was reported to be seat work (12), or a combination of seat work and teacher lecture (8) in 20 of the 23 studies, with three studies not reporting. The specific activity in which the children were engaged (e.g., structured vs. unstructured academic tasks; math vs. language arts) was reported in only two studies, precluding inclusion as a moderator.

Observation variables

Total duration. Observation duration was coded as (a) one to 10 minutes per observation session; (b) 11 to 20 minutes; or (c) 21 or more minutes.

Intervals. Within-observation intervals refer to the period of time spent observing before coding the behaviors of interest. Observation interval was coded as (a) one to 10 seconds of observation; (b) 11 or more seconds prior to coding, or (c) by incidence. The later category includes studies that coded each behavior as it changed rather than subdividing the total observation period into intervals.

Days of observation. Days of observation were coded as (a) one or two days of data recording; (b) three or four days of direct observation per subject, or (c) five or more days. For studies reporting a range of observation days, the minimum number of days for any subject was used.

Number of behaviors coded. The number of behaviors simultaneously coded by study observers was coded to retain groups of approximately equal size: (a) one to six, (b) 7-12, (c) 13-20, or (d) more than 21 behaviors.
**Number of observers.** The total number of observers collecting data for a study was not reported in nine of the 23 studies (39.1%), precluding inclusion in weighted regression analysis requiring listwise deletion of missing data (i.e., including this variable would have decreased by eight the number of studies analyzed, thus severely decreasing power and generalizability). This variable will be analyzed using the analog to ANOVA technique to determine any systematic relationship between the number of observers and obtained effect size.

**Definition of off-task.** Definition of off-task behavior refers to the minimum duration a child must be off-task to be coded as such. This category was coded as (a) partial interval, where time equal to less than a defined observational interval (as defined above) must pass before a child is coded on-/off-task (e.g., DuPaul & Rapport, 1993, define 15-s coding intervals, during which a child must be off-task for two or more consecutive seconds to be coded off-task for that interval); (b) whole interval, where the subject must be off-task the entire interval, or; (c) per incident, for studies coding each behavior change as it occurs. If a study coded children as on-task only when they were engaged appropriately during the entire interval, the definition of off-task was coded as partial – logically if the child were off-task at any point during an interval, he would not have been coded as on-task.

**Observation method.** Observation method was coded dichotomously. The *alternating* category includes studies observing ADHD and control children on each day in a multiple subsequent pattern. For example, the Classroom Observation Code used by Abikoff and colleagues (1977, 1980, 1984, 1985, & 2002) requires a four-minute observation of an ADHD child followed by four minutes observing a comparison child, repeated until 16 minutes of data is collected for each child. The *continuous* category includes studies observing one child for the
full duration reported. The total number of days and minutes each child is observed using either method is equal.

**Computation of Effect Sizes**

*Hedges' g*

Hedges’ *g* (1982) effect sizes were calculated to estimate the magnitude of differences in observed attention between children with ADHD and control children in classroom settings. Hedges’ *g* provides a correction to the standardized mean difference effect size to correct for the upward bias of studies with small sample size. Conceptually, an effect size is the amount of difference in standard deviation units between comparison groups (Lipsey & Wilson, 2001). An effect size of 1.0 indicates that the experimental group on average scored one standard deviation higher than the comparison group on the outcome in question. Effect sizes (ES) are categorized as small (ES ≤ 0.30), medium (ES ≈ 0.50), and large (ES ≥ 0.67) effects. These values are based on an analysis of over 300 published meta-analyses, and are used in lieu of those originally proposed by Cohen (1977). Means, standard deviations, and sample sizes were used to compute Hedges’ *g* for 17 studies (77%). An additional four studies (18%) provided sample size and *p*- or *t*-values needed for estimation, and one effect size was calculated using between-group difference and pooled standard deviation (Solanto et al., 2001). DuPaul, Ervin, Hook, and McGoey (1998) reported individual means for subjects, which were used for the calculation of group means and standard deviations. The Comprehensive Meta-Analysis software package was used to calculate effect sizes. Mean effect sizes were weighted by their inverse variance weight (*w*) during subsequent analyses to correct for imprecision associated with larger standard error, such that each study contributes in proportion to its sample size (Lipsey & Wilson, 2001).
Multiple effect sizes

Three studies reported data sufficient to calculate multiple effect sizes. Only one effect size was used for each study to meet the independence assumption (Lipsey & Wilson, 2001). Abikoff et al. (2002) reported separate descriptive statistics for males and females (N = 806 and 198, respectively), which were pooled based on recommendations by Lipsey and Wilson (2001). Zentall (1980) provided data for both off-task frequency and duration. Duration data were used to match the data format reported in the majority of studies. Jacob, O'Leary, and Rosenblad (1978) provided data for both formal and informal classroom settings. The formal classroom data was chosen because it consisted of teacher-led assignments typical of a normal classroom setting.

Nonsignificant results

Schecket and Schecket (1976) did not provide data sufficient to calculate effect size, but reported no significant between-group differences. An effect size of 0.00 was adopted for this study to minimize Type I error consistent with recommendations addressing the file drawer problem (i.e., publication bias favoring large, statistically significant effects; Rosenthal, 1995). This method is conservative in nature and may artificially decrease effect size estimates. Excluding such studies, however, may result in artificially small p-values and larger effect size estimates. The fail-safe N was computed to determine the potential bias of unpublished studies. This procedure estimates the number of studies with a 0.0 effect size that would be needed to decrease the overall mean effect size such that its confidence interval contained 0.0 (i.e., overall no significant differences between study groups). A rank correlation test was also used to further analyze the potential effects of publication bias (described below; Lipsey & Wilson, 2001).
Data Analysis

A four-tier data analytic strategy was adopted. In the first tier, Hedges’ $g$ effect sizes were calculated for the variable off-task in each study. Studies reporting on-task percentages were converted to off-task using the formula $100 – \text{on-task} \%$. Studies reporting number of intervals on-/off-task were converted to percentage off-task by dividing by the total number of intervals. An initial overall mean effect size was computed for fixed effects and tested for homogeneity using the $Q$ statistic to determine whether the amount of variance between studies could be attributed to random, study-level error variance (Lipsey & Wilson, 2001). A significant $Q$ rejects the assumption of homogeneity and requires further analysis.

Inverse variance weights ($w_i = 1/\text{SE}^2$) were calculated for each effect size in the second tier. These values cause studies with larger sample sizes, and hence less sampling error, to be weighted more heavily than studies with smaller sample sizes and larger standard errors. A fixed effects weighted regression approach using SPSS for Windows 12.0 was adopted. This approach assumes that additional variance is systematically related to measured study variables. It provides two measures of overall fit: $Q_R$, which reflects variance accounted for by the regression model ($p$ degrees of freedom, where $p$ equals the number of predictors); and $Q_E$, variance unaccounted for by either the model or random, study-level sampling error ($k – p – 1$ degrees of freedom, where $k$ equals the number of studies). Both statistics are distributed as chi-square.

Note: the $Q_B$ and $Q_W$ analogous to ANOVA technique reported in many meta-analytic reviews was not used for primary analyses because it inflates Type I error when used with several moderator variables – see Lipsey & Wilson, 2001 for discussion of this problem.
Corrected $B$-weight standard error for each moderator variable was calculated and each variable was tested against the $z$-distribution (Lipsey & Wilson, 2001).

In the third tier, a modified regression equation using the significant moderators from tier two was created to estimate the mean effect size that would likely have been obtained had all studies employed the best case combination of moderator variables found to significantly influence study results. The regression equation is solved by selecting the value of each moderator corresponding to empirically validated best practice (Lipsey & Wilson, 2001). To further explicate the results, original metric scores were calculated for each group to examine the difference in frequency of off-task behavior between children with ADHD and typically developing controls by adding the product of the overall weighted mean effect size and the control group standard deviation to the overall weighted mean of the control groups (Lipsey & Wilson, 2001). This procedure is similar to computing standard difference scores for raw data – it translates the effect size onto the control group distribution (e.g., Losier, McGrath, & Klein, 1996). Power analysis was conducted using GPower (Faul & Erdfelder, 1992) to determine the minimum number of subjects future studies will likely need to detect between-group differences in classroom attention given the best case effect size. Power analysis was used not as a theoretically suspect a posteriori analysis (i.e., there is no chance of a Type II error if obtained results are significant; Cohen, 1992), but as a first step for future research of the classroom attention of children with ADHD.

In the final tier, the 59 single case/case study design studies were analyzed and compared to the 23 between-group studies to determine the generalizability of the results obtained from the above procedures. Demographic and methodological variables and ADHD off-task rates for single case design studies are shown in Table 7. Three analyses were completed. Weighted
means for both groups of studies were compared using a t-test, power analysis, and area under the curve calculations. The Jacobson and Truax (1991) model of clinical significance informed this analysis, and computations were repeated after accounting for differences in the number of studies in each group. The second analysis involved the calculation of effect sizes for the six studies including a typically developing contrast group. Effect sizes could not be calculated for the vast majority (56 of 61) of these studies because they did not include a normal control group. In all six cases, peer group sample size was unreported and set equal to the number of children with ADHD to provide rough estimates of effect size. Obtained effect sizes were compared to the effect sizes predicted by the regression equation obtained in the between-group analysis. The third analysis compared between-group and single case studies using a series of ANOVAs for each of the significant moderators found in the between-group analyses. Bonferroni corrections for multiple comparisons were employed.
CHAPTER THREE: RESULTS

Tier I: Moderator-independent attentional differences

Publication Bias: The File Drawer Problem

A Fail-safe $N$ analysis was computed to determine the likelihood that missing studies would significantly influence the obtained mean effect sizes (Lipsey & Wilson, 2001; Rosenthal, 1991). Results indicate that an unlikely 1,829 studies would be needed to reduce the confidence interval of the mean effect size to include zero (i.e., result in no significant differences between off-task rates of children with ADHD and typically developing children in the classroom). A rank correlation test (Begg & Mazumdar, 1994) for publication bias was non-significant, Kendall's $\tau_b = 0.155, p = 0.15$. The Trim and Fill procedure suggests that zero studies are missing from the analysis based on expected symmetry when plotting effect sizes by the inverse of their standard errors (Duval & Tweedie, 2000). Collectively, these analyses suggest that the effect of publication bias is minimal in the current meta-analysis.

Off-task Comparisons

Off-task rates are shown in Table 2. Children with ADHD were off-task an average of 28.15% across studies (range = 5.11% to 83.0%), compared with 14.96% (range = 0.78% to 70%) for control children for the 20 studies reporting this data. All but two studies reported significant between-group differences (Cunningham & Siegel, 1987; Shecket & Shecket, 1976). At the group level, children with ADHD were more variable than control children across studies. The mean standard deviation across studies was significantly greater for ADHD groups than control groups, $t(16) = 4.76, p < .0005$. 
**Effect sizes**

Effect sizes (ES), standard error of effect size ($SE_{ES}$), and standard difference scores (SDS) are shown in Table 2. Mean weighted effect size for the 23 studies was 0.71 (95% CI = 0.62-0.79) with a range of 0.00 to 2.23. This result corresponds to a large effect based on Lipsey and Wilson (2001) criteria. The overall test of homogeneity suggests that there is more variance among the effect sizes than would be predicted by study-level error alone ($Q = 119.81$, $p < .0005$). A stem and leaf representation of the distribution of effect sizes is shown in Table 3.

**Tier II: Moderators of obtained effect size**

**Fixed Effects Model**

Analysis of moderator variables was conducted on 15 studies reporting data for all variables of interest. An ANOVA was conducted to determine whether there were systematic differences between studies reporting all data and studies with missing data. Results indicate no significant differences for effect size ($F = 0.66$, $p = .43$) or any moderator variables except classroom matching ($F = 7.34$, $p = .013$) and diagnostic method ($F = 9.05$, $p = .007$). Classroom matching, but not diagnostic method, was no longer significant after correcting for the number of comparisons. Studies with missing diagnostic data include all four studies diagnosing based solely on a referral for behavior problems, and two of the seven (29%) studies using single informant rating scales. When the behavior problem code was eliminated, no significant differences remained ($F = 0.351$, $p = .56$). Diagnostic method was therefore retained in the model, with the code for referral deleted.

A fixed effects weighted regression model was adopted to examine the influence of potential moderator variables on the observed variability in effect sizes. Results of the fixed
effects weighted regression reveal that the moderator variables described below predicted a large
amount of the variance in effect sizes (adjusted $R^2 = .91; Q_R = 91.93, p < .001; Q_E = 0.58, p =
ns$). $Q_R$ is analogous to the F-test for the $\chi^2$ distribution, wherein a significant $Q_R$ indicates that
the model predicts significant variability in the effect sizes (Lipsey & Wilson, 2001). A
nonsignificant $Q_E$ indicates that only subject-level sampling error remains across effect sizes
after accounting for variability explained by the model (Lipsey & Wilson, 2001). The influence
of each study variable is described below.

Sample

Total N. Sample size was reported in all 23 studies, and ranged from 16 to 1004 children,
with a median of 56 children. Sample size was used in the weighting of effect sizes and was
therefore not analyzed further.

Percent Male. Gender was reported in 20 of 23 studies, with males representing 84.6%
(range = 33% to 100%) of all subjects. The proportion of male to female subjects was not a
significant moderator in the regression analysis ($z = 1.50, ns$).

Matching. ADHD and control children were matched on two or more variables in 14 of
the 23 studies (61%). One study matched only on age (Werry & Quay, 1969), and seven studies
did not report matching on any variables. Fourteen (61%) studies matched based on sex, seven
(30%) based on age, 10 (44%) by grade in school, and seven (30%) by classroom. Age/grade ($z
= -4.58, p < .01$) and classroom matching ($z = -2.45, p < .01$) were both significant predictors of
effect size in the regression. Matching on these variables was associated with smaller effect
sizes. Gender matching was not used in the final regression due to lack of variability (see above).
An analog to ANOVA was conducted to determine the potential influence of gender matching on
obtained effect sizes (Lipsey & Wilson, 2001). Results reveal that gender matching did not account for significant differences in effect sizes ($Q_B = 2.90, p = .09$; $Q_W = 19.22, p = .57$).

*Mean Age.* Mean age (8.36 years, range of means = 7.2 to 9.6 years) was reported for both ADHD and control groups in 20 of the 23 studies (87%). Age mean was not analyzed as a moderator due to the restricted range across studies.

**Diagnostics**

*Method.* Four general methods were used to diagnose subjects, including formal semi-structured diagnostic interviews (26%), multiple informant rating scales (i.e., parent and teacher; 22%), single informant rating scales (teacher only, 26%; parent only, 4%) and referral from teacher (13%) or pediatrician (4%) without subsequent reported rating scales/clinical interviews. One study did not report diagnostic method (Werry & Quay, 1969). Of the five studies using diagnostic clinical interviews with the parent, four (80%) also used rating scales from multiple informants, and one added a parent rating scale. Rating scales were used in a majority of studies (N=17; 74%) to select sample participants. Diagnostic method significantly predicted effect size in the model ($z = -2.43, p < .05$), with more stringent diagnostic procedures associated with smaller effect sizes.

*Moniker.* The ADHD/ADDH moniker was used in eight of 23 (35%) studies, 12 of 23 (52%) studies used the term hyperactive or hyperkinetic, and two studies described their sample as behavioral problem children (9%). Diagnostic moniker significantly predicted effect size ($z = 3.19, p < .01$) after eliminating the behavior problem code due to missing data (see above), with
studies using children with ADDH/ADHD reporting larger effect sizes than studies of hyperactive children⁴.

Classroom type

Children were observed in their regular classroom environments (78% of studies) or simulated classrooms (22%). Classroom type was linearly related to obtained effect size ($z = -1.69, p < .05$), with regular classroom settings associated with larger effect sizes than simulated classrooms.

Observation Variables

Total duration. Children's behavior was coded by trained observers for periods ranging from 10 to 90 minutes, with a median of 16 minutes. Total duration of the observation interval (observation length) was reported in 22 of the 23 studies. Total daily duration was not linearly related to observed differences between time on-task for children with ADHD and their peers ($z = 1.07, ns$).

Intervals. Within each session, observation intervals lasted between two and 20 seconds before recording, with a median of 15 seconds across studies. Three studies (13%) coded by incident and did not divide the observation session into intervals. Interval duration was linearly related to obtained effect size ($z = 2.66, p < .01$), with longer intervals before coding associated with larger effect sizes.

---

⁴ Diagnostic moniker and publication year were not entered into the model together because the high correlation between these variables violates the assumption of heteroscedasticity. Publication year was therefore omitted from the model. Rerunning the weighted regression model substituting publication year for diagnostic moniker did not change the significance of any variable or the direction of any variable's $B$-weight. Publication year was not a significant predictor, $B$-weight = 0.137, $p > .05$, in the alternate model.
Days of observation. Total sessions ranged from one to 10 days of observations, with a median of 3 days across studies (100% reporting). Days of observation was significantly and linearly related to observed group differences, with more days associated with larger effect sizes ($z = 4.36, p < .01$).

Number of behaviors coded. Behavioral ratings by trained observers ranged from simple recording schemes involving as few as one category (DuPaul & Rapport, 1993) to complex observations of as many as 96 (Skansgaard & Burns, 1998) distinct behaviors across the 15 studies reporting. Many of the observers (56% of studies) were simultaneously observing 10 or more behaviors. The mean number of behaviors observed was 14.5, with a median of 11 observations per session across the 22 studies reporting this data. The number of behaviors coded simultaneously was significantly related to effect size, with greater numbers of behaviors associated with larger effect size ($z = 2.94, p < .01$).

Number of observers. A mean of 3.71 observers were used in each of the 14 studies reporting this information. Number of observers was not reported in nine of the 23 studies and was therefore not included in the regression analysis, as weighted regression in SPSS requires listwise deletion of missing data and inclusion would significantly decrease power. An analog to ANOVA computed with available data suggests that the number of observers is not significantly related to obtained effect size ($Q_B = 0.96, p = .33; Q_W = 8.24, p = .77$).

Definition of off-task. Some studies required children to remain off-task for an entire defined interval to be coded off-task, some coded off-task only if it occurred during part of a predefined interval, and others coded per incident without predefined intervals (nine, eight, and two studies, respectively, with four studies not reporting). The length of time a child must be off-
task to be coded as such was not significantly related to obtained effect size differences between children with and without ADHD ($z = 1.33$, $ns$)

*Observation method.* Eight studies used a continuous observation schema, whereas 14 alternated coding between experimental and control children within the same observation period. One study did not report observation method. Observation method predicted the magnitude of obtained effects, with continuous observation associated with smaller effect sizes ($z = -5.94$, $p < .01$).

**Tier III: Best Case Estimation and Original Metric**

*Best Case Estimation*

Values corresponding to age/grade and classroom matching, regular classroom settings, longer observation intervals, use of the ADHD moniker based on semi-structured clinical interviews and multiple informants, fewer behaviors simultaneously coded, continuous observation schemes, and greater total days of observation were selected based on best practice for each significant predictor in the regression equation (Lipsey & Wilson, 2001). Solving the regression equation using these values and corresponding $B$-weights (Table 4) suggests that an effect size of 1.40 would be expected on average for studies employing this combination of observational and diagnostic methodology, sample and classroom characteristics.

*Power Analysis for Future Studies*

GPower (Faul & Erdfelder, 1992) was used to determine the minimum sample size needed by future researchers to detect significant between-group differences in observed classroom attention of children with ADHD compared to their peers. With the best case effect
size of 1.4, an alpha of .05, and power of .80 as recommended by Cohen (1992), eight total subjects (four per group) are needed to detect between-group differences.

Original Metric

Across studies, typically developing children were off-task an average of 14.96% (SD = 16.47). An effect size of 1.40, based on the above estimation of the expected average effect size for research employing the most rigorous study design, corresponds to a 38.02% off-task average for children with ADHD. In other words, one would expect typically developing children to be on-task an average of 85.04% of the time, compared to 61.98% for children with ADHD. A graphical representation of this relationship is depicted in Figure 1.

Tier IV: Single Case Design Studies

Mean Differences

A total of 1,174 and 496 children with ADHD were observed in between-group and single case design studies (SCD), respectively. Children with ADHD were off-task a weighted average of 28.15% (SD = 18.28) in the 23 between-group studies, compared to 49.34% (SD = 16.41) in the 59 single case design studies. This difference was significant, \( t(36) = -4.85, p < .0005 \). The mean SCD off-task rate corresponds to 1.16 standard deviations above the between-group mean, calculated as the difference between means divided by the between-group standard deviation. Power analysis suggests that only 26 cases are needed to reject the null hypothesis given a difference of this magnitude, for an alpha of .05 and power set at .80 (Cohen, 1992). Because statistical significance is dependent on sample size, the results of the \( t \)-test may not inform clinical significance (Jacobson & Truax, 1991). Therefore, the number of SCD cases expected to equal or exceed 1.16 above the between-group mean was calculated to determine the
likelihood that the SCD studies come from a different population than the between-group studies. With an average between-group sample size of 51.04, and a normal distribution assumed, one would expect 6.28 subjects to meet or exceed 1.16 standard deviations above the mean, calculated as 51.04 times the area under curve (AUC = .1230). Multiplying the average number of SCD subjects by the area at or beyond the SCD mean (AUC = .5000; 8.13 x 0.5 = 4.08) suggests that fewer subjects than expected actually score at or above 1.16 SD above the between-group mean.

*Effect Size Comparisons*

Obtained and predicted effect sizes for the six SCD studies are displayed in Table 5. Obtained effect sizes (M = 3.53, SD = 1.29) did not differ significantly from effect sizes predicted by the regression equation, \( t(9) = 1.67, p = .117, \text{ ns} \). It is noted that the small number of studies analyzed may have influenced the failure to find statistically significant differences.

*Moderator Comparison*

Seven Bonferroni-corrected ANOVAs were conducted to analyze potential differences between between-group and SCD studies. Results are shown in Table 6. The studies differed on three of the variables found in Tier II to significantly influence observed differences between children with ADHD and typically developing children. In the Tier II regression equation, observation method is associated with smaller effect sizes, whereas the number of behaviors coded and number of observation days predicted larger magnitude differences.
<table>
<thead>
<tr>
<th>Study</th>
<th>Total N</th>
<th>Age Mean</th>
<th>Male/Female (%) Male</th>
<th>Match Ctrl</th>
<th>Classroom Type</th>
<th>Observation Duration (min)</th>
<th>Observation Interval (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Werry &amp; Quay (1969)</td>
<td>21</td>
<td>8.92</td>
<td>21/0 (100)</td>
<td>Y</td>
<td>Simulated</td>
<td>15</td>
<td>20</td>
</tr>
<tr>
<td>Forness &amp; Esveldt (1975)</td>
<td>48</td>
<td>7.2</td>
<td>48/0 (100)</td>
<td>Y Y Y</td>
<td>Regular</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>Shecket &amp; Shecket (1976)</td>
<td>36</td>
<td>NR</td>
<td>NR</td>
<td></td>
<td>Regular</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>Abikoff et al. (1977)</td>
<td>120</td>
<td>8.17</td>
<td>112/8 (93.3)</td>
<td>Y Y</td>
<td>Regular</td>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td>Campbell et al. (1978)</td>
<td>31</td>
<td>7.67</td>
<td>26/5 (83.9)</td>
<td>Y Y</td>
<td>Regular</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>Jacob et al. (1978)</td>
<td>16</td>
<td>9.6</td>
<td>14/2 (87.5)</td>
<td>Y</td>
<td>Simulated</td>
<td>30</td>
<td>10</td>
</tr>
<tr>
<td>Klein &amp; Young (1979)</td>
<td>34</td>
<td>NR</td>
<td>34/0 (100)</td>
<td></td>
<td>Regular</td>
<td>90</td>
<td>10</td>
</tr>
<tr>
<td>Abikoff et al. (1980)</td>
<td>119</td>
<td>8.4</td>
<td>114/5 (95.8)</td>
<td>Y Y</td>
<td>Regular</td>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td>Zentall (1980)</td>
<td>62</td>
<td>7.3</td>
<td>62/0 (100)</td>
<td>Y Y Y Y</td>
<td>Regular</td>
<td>10</td>
<td>Not divided</td>
</tr>
<tr>
<td>Abikoff &amp; Gittelman (1984)</td>
<td>56</td>
<td>8.5</td>
<td>54/2 (96.4)</td>
<td>Y Y</td>
<td>Regular</td>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td>Abikoff &amp; Gittelman (1985)</td>
<td>56</td>
<td>8.17</td>
<td>54/2 (96.4)</td>
<td>Y Y</td>
<td>Regular</td>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td>Atkins et al. (1985)</td>
<td>47</td>
<td>9.14</td>
<td>34/13 (72.3)</td>
<td>Y Y Y</td>
<td>Regular</td>
<td>NR</td>
<td>2</td>
</tr>
<tr>
<td>Book &amp; Skeen (1987)</td>
<td>162</td>
<td>NR</td>
<td>NR</td>
<td>Y Y</td>
<td>Regular</td>
<td>45</td>
<td>Not divided</td>
</tr>
<tr>
<td>Cunningham &amp; Siegel (1987)</td>
<td>60</td>
<td>8.71</td>
<td>60/0 (100)</td>
<td></td>
<td>Simulated</td>
<td>15</td>
<td>5</td>
</tr>
<tr>
<td>Roberts (1990)</td>
<td>33</td>
<td>8.92</td>
<td>33/0 (100)</td>
<td></td>
<td>Simulated</td>
<td>15</td>
<td>Not divided</td>
</tr>
<tr>
<td>DuPaul &amp; Rapport (1993)</td>
<td>56</td>
<td>8.32</td>
<td>46/10 (82.1)</td>
<td></td>
<td>Regular</td>
<td>20</td>
<td>15</td>
</tr>
<tr>
<td>Lett &amp; Kamphaus (1997)</td>
<td>55</td>
<td>7.94</td>
<td>41/14 (74.5)</td>
<td></td>
<td>Regular</td>
<td>15</td>
<td>3 (27sec recording)</td>
</tr>
<tr>
<td>Nolan &amp; Gadow (1997)</td>
<td>68</td>
<td>8.8</td>
<td>62/6 (91.2)</td>
<td></td>
<td>Regular</td>
<td>30</td>
<td>10</td>
</tr>
<tr>
<td>DuPaul et al. (1998)</td>
<td>26</td>
<td>7.7</td>
<td>21/5 (80.8)</td>
<td>Y Y Y</td>
<td>Regular</td>
<td>18</td>
<td>15</td>
</tr>
<tr>
<td>Skansgaard &amp; Burns (1998)</td>
<td>24</td>
<td>NR</td>
<td>8/16 (33.3)</td>
<td>Y Y Y</td>
<td>Regular</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>Solanto et al. (2001)</td>
<td>112</td>
<td>8.45</td>
<td>96/16 (85.7)</td>
<td></td>
<td>Regular</td>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td>Abikoff et al. (2002)</td>
<td>1004</td>
<td>8.4</td>
<td>806/198 (80.3)</td>
<td>Y Y</td>
<td>Regular</td>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td>Lauth &amp; Mackowiak (2004)</td>
<td>110</td>
<td>8.47</td>
<td>76/34 (69.1)</td>
<td>Y Y Y Y</td>
<td>Regular</td>
<td>45</td>
<td>NR</td>
</tr>
<tr>
<td>Study</td>
<td>Obs/Day</td>
<td>Number of Bx Coded</td>
<td>Number of Observers</td>
<td>Definition of Off-task</td>
<td>Diagnostic Method</td>
<td>Diagnostic Moniker</td>
<td>Observation Method</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-----------</td>
<td>--------------------</td>
<td>---------------------</td>
<td>------------------------</td>
<td>------------------</td>
<td>-------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Werry &amp; Quay (1969)</td>
<td>3</td>
<td>17</td>
<td>6</td>
<td>Partial &gt;5s</td>
<td>NR</td>
<td>Bx Prob</td>
<td>Continuous</td>
</tr>
<tr>
<td>Forness &amp; Esveldt (1975)</td>
<td>6</td>
<td>5</td>
<td>6</td>
<td>NR</td>
<td>TR</td>
<td>Bx Prob</td>
<td>Alternating</td>
</tr>
<tr>
<td>Shecket &amp; Shecket (1976)</td>
<td>4</td>
<td>19</td>
<td>1</td>
<td>NR</td>
<td>TR</td>
<td>Hyperactive</td>
<td>Alternating</td>
</tr>
<tr>
<td>Abikoff et al. (1977)</td>
<td>3</td>
<td>13</td>
<td>5</td>
<td>Whole &gt;15s</td>
<td>RS-M</td>
<td>Hyperactive</td>
<td>Alternating</td>
</tr>
<tr>
<td>Campbell et al. (1978)</td>
<td>2</td>
<td>5</td>
<td>NR</td>
<td>Whole &gt;15s</td>
<td>PR</td>
<td>Hyperactive</td>
<td>NR</td>
</tr>
<tr>
<td>Jacob et al. (1978)</td>
<td>10</td>
<td>6</td>
<td>NR</td>
<td>Partial &gt;10s</td>
<td>RS-T</td>
<td>Hyperactive</td>
<td>Alternating</td>
</tr>
<tr>
<td>Klein &amp; Young (1979)</td>
<td>2</td>
<td>17</td>
<td>2</td>
<td>Whole &gt;15s</td>
<td>RS-T</td>
<td>Hyperactive</td>
<td>Alternating</td>
</tr>
<tr>
<td>Abikoff et al. (1980)</td>
<td>3</td>
<td>12</td>
<td>4</td>
<td>Whole &gt;15s</td>
<td>RS-M</td>
<td>ADDH</td>
<td>Alternating</td>
</tr>
<tr>
<td>Zentall (1980)</td>
<td>5</td>
<td>6</td>
<td>3</td>
<td>Partial &gt;10s</td>
<td>RS-T</td>
<td>ADD</td>
<td>Alternating</td>
</tr>
<tr>
<td>Abikoff &amp; Gittelmann (1984)</td>
<td>3</td>
<td>12</td>
<td>7</td>
<td>Whole &gt;15s</td>
<td>RS-M</td>
<td>ADDH</td>
<td>Alternating</td>
</tr>
<tr>
<td>Abikoff &amp; Gittelmann (1985)</td>
<td>3</td>
<td>12</td>
<td>7</td>
<td>Whole &gt;15s</td>
<td>RS-M</td>
<td>ADDH</td>
<td>Alternating</td>
</tr>
<tr>
<td>Atkins et al. (1985)</td>
<td>7</td>
<td>32</td>
<td>2</td>
<td>Partial &gt;2s</td>
<td>SSI/RS-P</td>
<td>Hyperactive</td>
<td>Continuous</td>
</tr>
<tr>
<td>Book &amp; Skeen (1987)</td>
<td>1</td>
<td>4</td>
<td>NR</td>
<td>Partial &gt;10s</td>
<td>SSI/RS-P</td>
<td>Hyperactive</td>
<td>Continuous</td>
</tr>
<tr>
<td>Cunningham &amp; Siegel (1987)</td>
<td>1</td>
<td>16</td>
<td>2</td>
<td>Whole &gt;15s</td>
<td>SSI/RS-M</td>
<td>ADDH</td>
<td>Alternating</td>
</tr>
<tr>
<td>Roberts (1990)</td>
<td>1</td>
<td>6</td>
<td>NR</td>
<td>Whole &gt;15s</td>
<td>SSI/RS-P</td>
<td>ADDH</td>
<td>Continuous</td>
</tr>
<tr>
<td>Lett &amp; Kamphaus (1997)</td>
<td>1</td>
<td>13</td>
<td>NR</td>
<td>Partial &gt;2s</td>
<td>SSI/RS-M</td>
<td>ADDH</td>
<td>Continuous</td>
</tr>
<tr>
<td>Nolan &amp; Gadow (1997)</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>Whole &gt;15s</td>
<td>SSI/RS-M</td>
<td>ADDH</td>
<td>Alternating</td>
</tr>
<tr>
<td>Skansgaard &amp; Burns (1998)</td>
<td>4</td>
<td>96</td>
<td>2</td>
<td>Whole &gt;15s</td>
<td>SSI/RS-M</td>
<td>ADDH</td>
<td>Continuous</td>
</tr>
<tr>
<td>Solanto et al. (2001)</td>
<td>1</td>
<td>12</td>
<td>NR</td>
<td>Whole &gt;15s</td>
<td>SSI/RS-M</td>
<td>ADDH</td>
<td>Alternating</td>
</tr>
<tr>
<td>Abikoff et al. (2002)</td>
<td>3</td>
<td>12</td>
<td>NR</td>
<td>Whole &gt;15s</td>
<td>SSI/RS-M</td>
<td>ADDH</td>
<td>Alternating</td>
</tr>
<tr>
<td>Lauth &amp; Mackowiak (2004)</td>
<td>3</td>
<td>8</td>
<td>NR</td>
<td>Whole &gt;15s</td>
<td>SSI/RS-M</td>
<td>ADDH</td>
<td>Alternating</td>
</tr>
</tbody>
</table>
Notes. A = Age; Alternating = abab or similar method; Bx Prob = Referred for behavioral problems/hyperactivity; C = Classroom; Continuous =
Observing one child exclusively for entire observation period; Definition of Off-task = Proportion of interval needed to be coded off task; F = Females; G = Grade; M = Multiple informants; No of Bx Coded = Number of behaviors coded; NR = Not reported; Obs/Day = Number of observation days;
Observation Duration = Minutes of observation; Observation Interval = Seconds before recording; P = Parent rating scale Only; PR = Pediatrician referral;
RS = Rating scale(s); S = Sex; SSI = Semi-structured interview; T = Teacher rating scale only; TR = Teacher Referral.
Table 2: Mean Off-task Rates, Standard Difference Scores, and Effect Sizes in Children with ADHD

<table>
<thead>
<tr>
<th>Study</th>
<th>ADHD % Off-task M (SD)</th>
<th>Control % Off-task M (SD)</th>
<th>Std. Diff. Scores (%)</th>
<th>Hedges’ g Effect Sizes (Std. Error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Werry &amp; Quay (1969)</td>
<td>46.3 (12.8)</td>
<td>23 (15.4)</td>
<td>50.3</td>
<td>2.09 (0.53)</td>
</tr>
<tr>
<td>Forness &amp; Esveldt (1975)</td>
<td>47.0 (16.5)</td>
<td>34 (12.4)</td>
<td>27.7</td>
<td>0.88 (0.30)</td>
</tr>
<tr>
<td>Shecket &amp; Shecket (1976)</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>0.00</td>
</tr>
<tr>
<td>Abikoff et al. (1977)</td>
<td>13.1 (10.0)</td>
<td>2.1 (2.6)</td>
<td>84.2</td>
<td>1.50 (0.21)</td>
</tr>
<tr>
<td>Campbell et al. (1978)</td>
<td>16.73 (15.15)</td>
<td>12.41 (10.88)</td>
<td>25.8</td>
<td>0.32 (0.35)</td>
</tr>
<tr>
<td>Jacob et al. (1978)</td>
<td>15.8 (NR)</td>
<td>10.5 (NR)</td>
<td>33.3</td>
<td>1.41 (0.53)</td>
</tr>
<tr>
<td>Klein &amp; Young (1979)</td>
<td>39.8 (9.0)</td>
<td>26.6 (5.0)</td>
<td>33.1</td>
<td>1.78 (0.40)</td>
</tr>
<tr>
<td>Abikoff et al. (1980)</td>
<td>15.1 (23.4)</td>
<td>4.1 (7.8)</td>
<td>72.8</td>
<td>0.62 (0.19)</td>
</tr>
<tr>
<td>Zentall (1980)</td>
<td>15.0 (NR)</td>
<td>7.1 (NR)</td>
<td>52.2</td>
<td>0.45 (0.25)</td>
</tr>
<tr>
<td>Abikoff &amp; Gittelman (1984)</td>
<td>17.4 (12.3)</td>
<td>3.5 (6.6)</td>
<td>79.7</td>
<td>1.39 (0.29)</td>
</tr>
<tr>
<td>Abikoff &amp; Gittelman (1985)</td>
<td>15.7 (10.4)</td>
<td>2.5 (4.6)</td>
<td>84.1</td>
<td>1.71 (0.31)</td>
</tr>
<tr>
<td>Atkins et al. (1985)</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>0.59 (0.30)</td>
</tr>
<tr>
<td>Book &amp; Skeen (1987)</td>
<td>5.11 (4.82)</td>
<td>0.78 (1.47)</td>
<td>84.7</td>
<td>1.21 (0.17)</td>
</tr>
<tr>
<td>Cunningham &amp; Siegel (1987)</td>
<td>33.0 (NR)</td>
<td>26.4 (NR)</td>
<td>19.9</td>
<td>0.51 (0.26)</td>
</tr>
<tr>
<td>Roberts (1990)</td>
<td>39.5 (18.8)</td>
<td>12.9 (20.9)</td>
<td>67.3</td>
<td>1.31 (0.39)</td>
</tr>
<tr>
<td>DuPaul &amp; Rapport (1993)</td>
<td>44.26 (16.56)</td>
<td>19.72 (11.56)</td>
<td>55.4</td>
<td>1.66 (0.31)</td>
</tr>
<tr>
<td>Lett &amp; Kamphaus (1997)</td>
<td>18.3 (16.5)</td>
<td>12.7 (12.7)</td>
<td>30.6</td>
<td>0.36 (0.29)</td>
</tr>
<tr>
<td>Nolan &amp; Gadow (1997)</td>
<td>30.5 (15.9)</td>
<td>13.3 (8.3)</td>
<td>56.4</td>
<td>1.34 (0.27)</td>
</tr>
<tr>
<td>DuPaul et al. (1998)</td>
<td>33.0 (19.2)</td>
<td>9.5 (11.9)</td>
<td>71.2</td>
<td>1.31 (0.45)</td>
</tr>
<tr>
<td>Skansgaard &amp; Burns (1998)</td>
<td>23.8 (10.3)</td>
<td>4.8 (6.1)</td>
<td>79.8</td>
<td>2.23 (0.60)</td>
</tr>
<tr>
<td>Solanto et al. (2001)</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>0.58 (0.19)</td>
</tr>
<tr>
<td>Abikoff et al. (2002)</td>
<td>10.6 (24.0)</td>
<td>3.3 (13.2)</td>
<td>68.8</td>
<td>0.38 (0.06)</td>
</tr>
<tr>
<td>Lauth &amp; Mackowiak (2004)</td>
<td>83.0 (12.0)</td>
<td>70.0 (13.0)</td>
<td>15.7</td>
<td>1.03 (0.20)</td>
</tr>
</tbody>
</table>

Column M (SD) = 28.15 (18.28) 14.96 (16.47) 54.65 (23.71) 0.71 (0.04)
Notes: 
1. Effect size calculated using N = 47, $t = 2.01$;  
2. Effect size calculated using N = 60, $p = .052$;  
3. Effect size calculated using N = 16, $p = .01$;  
4. Effect size set at zero – insufficient data to calculate effect size of nonsignificant differences;  
5. Effect size calculated using differences in means (.007), common SD (.12), and N = 112;  
6. Standard error of effect sizes; SD = Mean of standard deviations for included studies.
Table 3: Stem and Leaf Display of 23 Hedges g Effect Sizes

<table>
<thead>
<tr>
<th>Stem</th>
<th>Leaf</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2</td>
<td>3</td>
</tr>
<tr>
<td>2.1</td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>9</td>
</tr>
<tr>
<td>1.9</td>
<td></td>
</tr>
<tr>
<td>1.8</td>
<td></td>
</tr>
<tr>
<td>1.7</td>
<td>1, 8</td>
</tr>
<tr>
<td>1.6</td>
<td>6</td>
</tr>
<tr>
<td>1.5</td>
<td>0</td>
</tr>
<tr>
<td>1.4</td>
<td>1</td>
</tr>
<tr>
<td>1.3</td>
<td>1, 1, 4, 9</td>
</tr>
<tr>
<td>1.2</td>
<td>1</td>
</tr>
<tr>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>3</td>
</tr>
<tr>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>8</td>
</tr>
<tr>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>2</td>
</tr>
<tr>
<td>0.5</td>
<td>1, 8, 9</td>
</tr>
<tr>
<td>0.4</td>
<td>5</td>
</tr>
<tr>
<td>0.3</td>
<td>2, 6, 8</td>
</tr>
<tr>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td>0</td>
</tr>
</tbody>
</table>

Note: A stem-and-leaf plot of classroom attentional differences. Effect sizes are arranged by place value such that digits in the left column represent the stem, while digits in the right columns represent the leaf.

For example, the effect sizes of .32, .36, and .38 are represented as .3 in the left column, and 2, 6, and 8 in the right columns, respectively.
Table 4: Inverse variance weighted regression results

<table>
<thead>
<tr>
<th>Model</th>
<th>$\chi^2$</th>
<th>df</th>
<th>$\chi^2$ significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>QR</td>
<td>91.93</td>
<td>13</td>
<td>$p &lt; .001$</td>
</tr>
<tr>
<td>QE</td>
<td>0.58</td>
<td>1</td>
<td>ns</td>
</tr>
</tbody>
</table>

$R^2_{\text{Model}} = .997$

Adjusted $R^2_{\text{Model}} = .912$

<table>
<thead>
<tr>
<th>Variable</th>
<th>$B$-weight</th>
<th>SE$_B$</th>
<th>Z-score</th>
<th>$Z_{\text{significance}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age/Grade Matching</td>
<td>-1.95042</td>
<td>0.426101</td>
<td>-4.57736</td>
<td>$p&lt;.01$</td>
</tr>
<tr>
<td>Classroom Matching</td>
<td>-2.02899</td>
<td>0.828252</td>
<td>-2.44973</td>
<td>$p&lt;.05$</td>
</tr>
<tr>
<td>Percent Male</td>
<td>0.38487</td>
<td>0.255887</td>
<td>1.504063</td>
<td>ns</td>
</tr>
<tr>
<td>Observation Duration (min.)</td>
<td>0.686629</td>
<td>0.640711</td>
<td>1.071667</td>
<td>ns</td>
</tr>
<tr>
<td>Observation Interval (sec.)</td>
<td>1.212349</td>
<td>0.455464</td>
<td>2.661788</td>
<td>$p&lt;.01$</td>
</tr>
<tr>
<td>Classroom Type</td>
<td>-1.69</td>
<td>0.396484</td>
<td>-4.26247</td>
<td>$p&lt;.01$</td>
</tr>
<tr>
<td>Definition Off-task</td>
<td>0.415595</td>
<td>0.311485</td>
<td>1.33424</td>
<td>ns</td>
</tr>
<tr>
<td>Diagnostic Moniker</td>
<td>0.94618</td>
<td>0.296885</td>
<td>3.187027</td>
<td>$p&lt;.01$</td>
</tr>
<tr>
<td>Observation Method</td>
<td>-2.4322</td>
<td>0.409737</td>
<td>-5.93601</td>
<td>$p&lt;.01$</td>
</tr>
<tr>
<td>Number of Behaviors Coded</td>
<td>0.857616</td>
<td>0.291726</td>
<td>2.939799</td>
<td>$p&lt;.01$</td>
</tr>
<tr>
<td>Days of Observations</td>
<td>1.844816</td>
<td>0.423337</td>
<td>4.357798</td>
<td>$p&lt;.01$</td>
</tr>
<tr>
<td>Diagnostic Method</td>
<td>-0.78683</td>
<td>0.30749</td>
<td>-2.55886</td>
<td>$p&lt;.05$</td>
</tr>
<tr>
<td>(Constant)</td>
<td>0.217208</td>
<td>1.56588</td>
<td>0.138713</td>
<td>ns</td>
</tr>
</tbody>
</table>

Note: $B$-weight = unstandardized regression weight; SE$_B$ = standard error of the regression weight.
Table 5: Predicted and obtained effect sizes for single case design studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Predicted ES</th>
<th>Obtained ES (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ward &amp; Baker (1969)</td>
<td>3.54</td>
<td>3.16 (1.00)</td>
</tr>
<tr>
<td>Glynn &amp; Quinnell (1972)</td>
<td>2.74</td>
<td>5.81 (1.80)</td>
</tr>
<tr>
<td>Walker &amp; Hops (1976)</td>
<td>1.93</td>
<td>3.35 (0.76)</td>
</tr>
<tr>
<td>Pelham et al. (1980)</td>
<td>3.28</td>
<td>1.82 (0.73)</td>
</tr>
<tr>
<td>Kraemer (1994) D</td>
<td>0.80</td>
<td>3.52 (0.64)</td>
</tr>
<tr>
<td>Fabiano &amp; Pelham (2003)</td>
<td>2.20</td>
<td>3.52 (1.37)</td>
</tr>
<tr>
<td><strong>Mean (SD)</strong></td>
<td><strong>2.42 (1.00)</strong></td>
<td><strong>3.53 (1.29)</strong></td>
</tr>
</tbody>
</table>

Note: Comparison group sample size not reported in these studies. Obtained effect sizes estimated by assuming an equal number of subjects per group. D = Dissertation; ES = Effect size; SD = Standard deviations of means.
Table 6: Between group and single case design study moderator variable differences

<table>
<thead>
<tr>
<th>Moderator</th>
<th>F-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observation Interval (seconds)</td>
<td>0.34</td>
<td>ns</td>
</tr>
<tr>
<td>Diagnostic Moniker</td>
<td>5.18</td>
<td>ns</td>
</tr>
<tr>
<td>Observation Method</td>
<td>9.39</td>
<td>.021</td>
</tr>
<tr>
<td>Number of Behaviors Coded</td>
<td>17.31</td>
<td>.004</td>
</tr>
<tr>
<td>Observation Days</td>
<td>11.98</td>
<td>.007</td>
</tr>
<tr>
<td>Diagnostic Method</td>
<td>1.84</td>
<td>ns</td>
</tr>
<tr>
<td>Classroom Type</td>
<td>5.08</td>
<td>ns</td>
</tr>
</tbody>
</table>

Note: p-value reported after Bonferroni correction (original p times number of comparisons).
<table>
<thead>
<tr>
<th>Study</th>
<th>Total N</th>
<th>Age Mean</th>
<th>Male/Female (% Male)</th>
<th>Classroom Type</th>
<th>Observation Duration (min)</th>
<th>Observation Interval (sec)</th>
<th>ADHD % Off-task M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quay et al. (1967)</td>
<td>5</td>
<td>7.52</td>
<td>100%</td>
<td>Regular</td>
<td>12.5</td>
<td>10</td>
<td>62% (NR)</td>
</tr>
<tr>
<td>Hall et al. (1968)</td>
<td>6</td>
<td>NR</td>
<td>67%</td>
<td>Regular</td>
<td>30</td>
<td>10</td>
<td>60.33% (15.17)</td>
</tr>
<tr>
<td>Ward &amp; Baker (1969)</td>
<td>8</td>
<td>NR</td>
<td>75%</td>
<td>Regular</td>
<td>15</td>
<td>20</td>
<td>74%</td>
</tr>
<tr>
<td>Buckley &amp; Walker (1971)</td>
<td>44</td>
<td>NR</td>
<td>88.6%</td>
<td>Regular</td>
<td>6</td>
<td>15</td>
<td>55% (NR)</td>
</tr>
<tr>
<td>Glynn &amp; Quinnell (1971)</td>
<td>6</td>
<td>NR</td>
<td>NR</td>
<td>Regular</td>
<td>35</td>
<td>10</td>
<td>53.67% (3.79)</td>
</tr>
<tr>
<td>Axelrod et al. (1972)</td>
<td>1</td>
<td>NR</td>
<td>100%</td>
<td>Regular</td>
<td>120 (2 hrs)</td>
<td>120 (2 min)</td>
<td>40% (NR)</td>
</tr>
<tr>
<td>Glynn &amp; Thomas (1974)</td>
<td>9</td>
<td>NR</td>
<td>89%</td>
<td>Regular</td>
<td>50</td>
<td>10</td>
<td>51.4%</td>
</tr>
<tr>
<td>McCullough et al. (1974)</td>
<td>1</td>
<td>6</td>
<td>100%</td>
<td>Regular</td>
<td>120 (2 hrs)</td>
<td>900 (15 min)</td>
<td>70.4% (8.17)</td>
</tr>
<tr>
<td>Walker et al. (1976)</td>
<td>10</td>
<td>NR</td>
<td>90%</td>
<td>Regular</td>
<td>12</td>
<td>15</td>
<td>64% (NR)</td>
</tr>
<tr>
<td>Walker &amp; Hops (1976)</td>
<td>24</td>
<td>NR</td>
<td>71%</td>
<td>Regular</td>
<td>9</td>
<td>6</td>
<td>72.4% (5.85)</td>
</tr>
<tr>
<td>Hay et al. (1977)</td>
<td>10</td>
<td>NR</td>
<td>100%</td>
<td>Regular</td>
<td>10</td>
<td>10</td>
<td>46.30% (12.76)</td>
</tr>
<tr>
<td>Marholin &amp; Steinman (1977)</td>
<td>8</td>
<td>NR</td>
<td>50%</td>
<td>Regular</td>
<td>15</td>
<td>10</td>
<td>67%</td>
</tr>
<tr>
<td>Walker (1977) D</td>
<td>12</td>
<td>NR</td>
<td>83.3%</td>
<td>Regular</td>
<td>16</td>
<td>30</td>
<td>38.33% (10.96)</td>
</tr>
<tr>
<td>Epstein &amp; Goss (1978)</td>
<td>1</td>
<td>10</td>
<td>100%</td>
<td>Regular</td>
<td>30</td>
<td>Momentary 75% (NR)</td>
<td></td>
</tr>
<tr>
<td>O’Leary &amp; Pelham (1978)</td>
<td>7</td>
<td>8.67</td>
<td>NR</td>
<td>Regular</td>
<td>15</td>
<td>10</td>
<td>53.38% (30.40)</td>
</tr>
<tr>
<td>Rose (1978)</td>
<td>2</td>
<td>8</td>
<td>0%</td>
<td>Regular</td>
<td>30</td>
<td>NR</td>
<td>25% (13.36)</td>
</tr>
<tr>
<td>Loney et al. (1979)</td>
<td>24</td>
<td>NR</td>
<td>100%</td>
<td>Regular</td>
<td>NR</td>
<td>NR</td>
<td>39% (NR)</td>
</tr>
<tr>
<td>Friedling &amp; O’Leary (1979)</td>
<td>4</td>
<td>7.58</td>
<td>87.5%</td>
<td>Regular</td>
<td>10</td>
<td>15</td>
<td>44% (4.1%)</td>
</tr>
<tr>
<td>Barkley et al. (1980)</td>
<td>6</td>
<td>8.35</td>
<td>100%</td>
<td>Simulated</td>
<td>10</td>
<td>15</td>
<td>50.25% (9.18)</td>
</tr>
<tr>
<td>Cameron &amp; Robinson (1980)</td>
<td>3</td>
<td>7.8</td>
<td>66.7%</td>
<td>Regular (SE)</td>
<td>25</td>
<td>10</td>
<td>43.31% (14.97)</td>
</tr>
<tr>
<td>Pelham et al. (1980)</td>
<td>8</td>
<td>8.3</td>
<td>87.5%</td>
<td>Regular</td>
<td>15</td>
<td>10</td>
<td>43%</td>
</tr>
<tr>
<td>Rapport et al. (1980)</td>
<td>2</td>
<td>7.5</td>
<td>50%</td>
<td>Regular</td>
<td>60</td>
<td>25</td>
<td>70.8% (12.27)</td>
</tr>
<tr>
<td>Eastman &amp; Rasbury (1981)</td>
<td>11</td>
<td>NR</td>
<td>NR</td>
<td>Regular</td>
<td>15</td>
<td>10</td>
<td>33.0% (11.00)</td>
</tr>
<tr>
<td>Hallahan et al. (1981)</td>
<td>3</td>
<td>10.3</td>
<td>100%</td>
<td>Regular (SE)</td>
<td>20</td>
<td>2</td>
<td>74.07% (7.13)</td>
</tr>
<tr>
<td>Krause (1981) D</td>
<td>1</td>
<td>7</td>
<td>100%</td>
<td>Regular</td>
<td>20</td>
<td>10</td>
<td>27%</td>
</tr>
<tr>
<td>Roberts &amp; Nelson (1981)</td>
<td>3</td>
<td>NR</td>
<td>100%</td>
<td>Regular</td>
<td>30</td>
<td>10</td>
<td>41.53% (NA)</td>
</tr>
<tr>
<td>Gettinger (1982)</td>
<td>8</td>
<td>7.9</td>
<td>50%</td>
<td>Regular (SE)</td>
<td>30</td>
<td>10</td>
<td>82.1% (11.3)</td>
</tr>
<tr>
<td>Hallahan et al. (1982)</td>
<td>1</td>
<td>8</td>
<td>100%</td>
<td>Regular (SE)</td>
<td>20</td>
<td>4</td>
<td>54.38% (11.71)</td>
</tr>
<tr>
<td>Lloyd et al. (1982)</td>
<td>3</td>
<td>9.33</td>
<td>66.7%</td>
<td>Regular (SE)</td>
<td>NR</td>
<td>2</td>
<td>48.93% (28.03)</td>
</tr>
<tr>
<td>Study</td>
<td>Total N^1</td>
<td>Age Mean</td>
<td>Male/Female (% Male)</td>
<td>Classroom Type</td>
<td>Observation Duration (min)</td>
<td>Observation Interval (sec)</td>
<td>ADHD % Off-task</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-----------</td>
<td>----------</td>
<td>----------------------</td>
<td>----------------</td>
<td>---------------------------</td>
<td>---------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Rapport et al. (1982)</td>
<td>2</td>
<td>7.5</td>
<td>100%</td>
<td>Regular</td>
<td>20</td>
<td>10</td>
<td>58.5% (8.19)^2</td>
</tr>
<tr>
<td>Rhode et al. (1983)</td>
<td>6</td>
<td>8.7</td>
<td>NA</td>
<td>Regular</td>
<td>15</td>
<td>10</td>
<td>61.83% (10.65)</td>
</tr>
<tr>
<td>Shepp &amp; Jensen (1983)</td>
<td>1</td>
<td>7</td>
<td>100%</td>
<td>Regular (SE)</td>
<td>60</td>
<td>10</td>
<td>38.0% (2.83)^2</td>
</tr>
<tr>
<td>Rooney et al. (1984)</td>
<td>4</td>
<td>NR</td>
<td>50%</td>
<td>Regular</td>
<td>20</td>
<td>2</td>
<td>86.95% (11.99)^2</td>
</tr>
<tr>
<td>Rosen et al. (1984)</td>
<td>23</td>
<td>NR</td>
<td>73.9%</td>
<td>Regular (SE)</td>
<td>30</td>
<td>10</td>
<td>20.26% (NR)^2</td>
</tr>
<tr>
<td>Rapport et al. (1985)</td>
<td>12</td>
<td>NR</td>
<td>100%</td>
<td>Regular</td>
<td>20</td>
<td>10</td>
<td>47% (NR)^2</td>
</tr>
<tr>
<td>Burlton-Bennet et al. (1987)</td>
<td>1</td>
<td>6</td>
<td>100%</td>
<td>Regular</td>
<td>30</td>
<td>10</td>
<td>74% (NR)</td>
</tr>
<tr>
<td>Rapport et al. (1987)</td>
<td>42</td>
<td>8.5</td>
<td>88.1%</td>
<td>Regular</td>
<td>20</td>
<td>15</td>
<td>46.5% (18.80)</td>
</tr>
<tr>
<td>Walker (1989) D</td>
<td>6</td>
<td>NR</td>
<td>100%</td>
<td>Regular</td>
<td>20</td>
<td>15</td>
<td>71.25% (9.01)</td>
</tr>
<tr>
<td>Szumlja (1990) D</td>
<td>8</td>
<td>NR</td>
<td>75%</td>
<td>Regular</td>
<td>30</td>
<td>10</td>
<td>37.04% (NR)</td>
</tr>
<tr>
<td>Bloomquist et al. (1991)</td>
<td>13</td>
<td>8.81</td>
<td>69.23%</td>
<td>Regular</td>
<td>10</td>
<td>10</td>
<td>49.4% (10.0)</td>
</tr>
<tr>
<td>DuPaul et al. (1992)</td>
<td>1</td>
<td>7</td>
<td>100%</td>
<td>Regular (SE)</td>
<td>15</td>
<td>30</td>
<td>58.5% (NR)</td>
</tr>
<tr>
<td>Paniagua (1992)</td>
<td>1</td>
<td>10</td>
<td>100%</td>
<td>Regular</td>
<td>10</td>
<td>30</td>
<td>88.6% (12.19)^2</td>
</tr>
<tr>
<td>DuPaul &amp; Henningson (1993)</td>
<td>1</td>
<td>7</td>
<td>100%</td>
<td>Regular</td>
<td>10</td>
<td>NR</td>
<td>61%</td>
</tr>
<tr>
<td>Kraemer (1994) D</td>
<td>12</td>
<td>8.33</td>
<td>NR</td>
<td>Regular</td>
<td>10</td>
<td>20</td>
<td>52.00% (8.88)</td>
</tr>
<tr>
<td>Rapport et al. (1994)</td>
<td>76</td>
<td>8.51</td>
<td>86.8%</td>
<td>Regular</td>
<td>20</td>
<td>15</td>
<td>44% (NR)^2</td>
</tr>
<tr>
<td>Kelley &amp; McCain (1995)</td>
<td>3</td>
<td>7.0</td>
<td>67%</td>
<td>Simulated</td>
<td>45</td>
<td>15</td>
<td>53.67%</td>
</tr>
<tr>
<td>Turner (1996) D</td>
<td>6</td>
<td>8.99</td>
<td>33.3%</td>
<td>Regular</td>
<td>10</td>
<td>10</td>
<td>46% (NR)^2</td>
</tr>
<tr>
<td>Matheson (1997) D</td>
<td>3</td>
<td>10</td>
<td>100%</td>
<td>Regular</td>
<td>15</td>
<td>30</td>
<td>66% (range: 44-90%)</td>
</tr>
<tr>
<td>Sams (1999) D</td>
<td>4</td>
<td>NR</td>
<td>100%</td>
<td>Regular</td>
<td>20</td>
<td>10</td>
<td>38% (NR)^2</td>
</tr>
<tr>
<td>Cloward (2000) D</td>
<td>8</td>
<td>NR</td>
<td>63%</td>
<td>Regular</td>
<td>10</td>
<td>6</td>
<td>26.16% (7.18)</td>
</tr>
<tr>
<td>Swenson et al. (2000)</td>
<td>1</td>
<td>12</td>
<td>100%</td>
<td>Regular (SE)</td>
<td>60</td>
<td>NA</td>
<td>41.33% (NR)</td>
</tr>
<tr>
<td>Roberts et al. (2001)</td>
<td>3</td>
<td>NR</td>
<td>100%</td>
<td>Regular</td>
<td>15</td>
<td>15</td>
<td>70.67% (NR)</td>
</tr>
<tr>
<td>Northup &amp; Gulley (2001)</td>
<td>1</td>
<td>NR</td>
<td>NR</td>
<td>Regular</td>
<td>5</td>
<td>10</td>
<td>10% (NR)</td>
</tr>
<tr>
<td>Austin (2003) D</td>
<td>1</td>
<td>8.33</td>
<td>100%</td>
<td>Regular</td>
<td>20</td>
<td>15</td>
<td>56% (NR)</td>
</tr>
<tr>
<td>Fabiano &amp; Pelham (2003)</td>
<td>1</td>
<td>9.92</td>
<td>100%</td>
<td>Regular</td>
<td>50</td>
<td>1</td>
<td>33.22% (9.82)</td>
</tr>
<tr>
<td>Lorah (2003) D</td>
<td>3</td>
<td>9.25</td>
<td>100%</td>
<td>Regular</td>
<td>15</td>
<td>15</td>
<td>27% (NR)</td>
</tr>
<tr>
<td>Dobrinski (2004) D</td>
<td>4</td>
<td>8</td>
<td>75%</td>
<td>Regular</td>
<td>20</td>
<td>10</td>
<td>56.38% (15.48)</td>
</tr>
<tr>
<td>Trolinder et al. (2004)</td>
<td>2</td>
<td>2</td>
<td>50%</td>
<td>Regular</td>
<td>20</td>
<td>10</td>
<td>47.15% (14.24)</td>
</tr>
<tr>
<td>Clarfield &amp; Stoner (2005)</td>
<td>1</td>
<td>7</td>
<td>100%</td>
<td>Regular</td>
<td>10</td>
<td>15</td>
<td>49% (NR)</td>
</tr>
</tbody>
</table>

40
<table>
<thead>
<tr>
<th>Study</th>
<th>Obs/Day of Bx Coded</th>
<th>Number of Observers</th>
<th>Definition of Off-task</th>
<th>Diagnostic Method</th>
<th>Diagnostic Moniker</th>
<th>Observation Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quay et al. (1967)</td>
<td>10</td>
<td>1</td>
<td>Partial (any)</td>
<td>TR</td>
<td>Hyperactive</td>
<td>Alternating</td>
</tr>
<tr>
<td>Hall et al. (1968)</td>
<td>7</td>
<td>3</td>
<td>Partial &gt;5s</td>
<td>TR</td>
<td>Bx Prob</td>
<td>Continuous</td>
</tr>
<tr>
<td>Ward &amp; Baker (1969)</td>
<td>20</td>
<td>12</td>
<td>Partial (any)</td>
<td>TR</td>
<td>Bx Prob</td>
<td>Continuous</td>
</tr>
<tr>
<td>Buckley &amp; Walker (1971)</td>
<td>NR</td>
<td>21</td>
<td>NR</td>
<td>T RS</td>
<td>Bx Prob</td>
<td>Continuous</td>
</tr>
<tr>
<td>Glynn &amp; Quinnell (1971)</td>
<td>10</td>
<td>8</td>
<td>Partial</td>
<td>TR</td>
<td>Bx Prob</td>
<td>Alternating</td>
</tr>
<tr>
<td>Axelrod et al. (1972)</td>
<td>5</td>
<td>1</td>
<td>NR</td>
<td>Glance</td>
<td>Bx Prob</td>
<td>Continuous</td>
</tr>
<tr>
<td>Glynn &amp; Thomas (1974)</td>
<td>10</td>
<td>8</td>
<td>Partial</td>
<td>TR</td>
<td>Bx Prob</td>
<td>Alternating</td>
</tr>
<tr>
<td>McCullough et al. (1974)</td>
<td>5</td>
<td>NR</td>
<td>Partial (any)</td>
<td>TR</td>
<td>Bx Prob</td>
<td>Continuous</td>
</tr>
<tr>
<td>Walker et al. (1976)</td>
<td>10</td>
<td>11</td>
<td>NR</td>
<td>T RS</td>
<td>Bx Prob</td>
<td>Alternating</td>
</tr>
<tr>
<td>Walker &amp; Hops (1976)</td>
<td>2</td>
<td>19</td>
<td>NR</td>
<td>TR</td>
<td>Bx Prob</td>
<td>Alternating</td>
</tr>
<tr>
<td>Hay et al. (1977)</td>
<td>7</td>
<td>2</td>
<td>Glance</td>
<td>TR</td>
<td>Bx Prob</td>
<td>Continuous</td>
</tr>
<tr>
<td>Marholin &amp; Steinman (1977)</td>
<td>7</td>
<td>3</td>
<td>Partial &gt;2s</td>
<td>TR</td>
<td>Bx Prob</td>
<td>Alternating</td>
</tr>
<tr>
<td>Walker (1977) D</td>
<td>10</td>
<td>1</td>
<td>NR</td>
<td>PR</td>
<td>Hyperactive</td>
<td>Continuous</td>
</tr>
<tr>
<td>Epstein &amp; Goss (1978)</td>
<td>3</td>
<td>3</td>
<td>Glance</td>
<td>TR</td>
<td>Bx Prob</td>
<td>Continuous</td>
</tr>
<tr>
<td>O’Leary &amp; Pelham (1978)</td>
<td>3</td>
<td>NR</td>
<td>NR</td>
<td>PR</td>
<td>Hyperkinetic</td>
<td>Alternating</td>
</tr>
<tr>
<td>Rose (1978)</td>
<td>2</td>
<td>3</td>
<td>Incidence &gt;5s</td>
<td>PR</td>
<td>Hyperactive</td>
<td>Incidence</td>
</tr>
<tr>
<td>Loney et al. (1979)</td>
<td>1</td>
<td>2</td>
<td>NR</td>
<td>P&amp;T RS</td>
<td>Hyperkinetic</td>
<td>NR</td>
</tr>
<tr>
<td>Friedling &amp; O’Leary (1979)</td>
<td>10</td>
<td>2</td>
<td>Partial</td>
<td>P&amp;T RS</td>
<td>Hyperactive</td>
<td>Alternating</td>
</tr>
<tr>
<td>Barkley et al. (1980)</td>
<td>4</td>
<td>2</td>
<td>Partial &gt;5s</td>
<td>P&amp;T RS</td>
<td>Hyperactive</td>
<td>Continuous</td>
</tr>
<tr>
<td>Cameron &amp; Robinson (1980)</td>
<td>13</td>
<td>1</td>
<td>Partial (any)</td>
<td>T RS</td>
<td>Hyperactive</td>
<td>Alternating</td>
</tr>
<tr>
<td>Pelham et al. (1980)</td>
<td>6</td>
<td>1</td>
<td>NR</td>
<td>P&amp;T RS</td>
<td>Hyperactive</td>
<td>Alternating</td>
</tr>
<tr>
<td>Rapport et al. (1980)</td>
<td>7</td>
<td>NR</td>
<td>Partial &gt;2 s</td>
<td>PR</td>
<td>Impulsive</td>
<td>NR</td>
</tr>
<tr>
<td>Eastman &amp; Rasbury (1981)</td>
<td>5</td>
<td>2</td>
<td>NR</td>
<td>T RS</td>
<td>Hyperkinetic</td>
<td>Continuous</td>
</tr>
<tr>
<td>Hallahan et al. (1981)</td>
<td>5</td>
<td>1</td>
<td>Glance</td>
<td>TR</td>
<td>Att Px</td>
<td>Alternating</td>
</tr>
<tr>
<td>Krause (1981) D</td>
<td>6</td>
<td>8</td>
<td>NA</td>
<td>NA</td>
<td>Hyperactive</td>
<td>Continuous</td>
</tr>
<tr>
<td>Roberts &amp; Nelson (1981)</td>
<td>7</td>
<td>2</td>
<td>Glance</td>
<td>T RS</td>
<td>Bx Prob</td>
<td>Continuous</td>
</tr>
<tr>
<td>Gettinger (1982)</td>
<td>18</td>
<td>3</td>
<td>NR</td>
<td>NR</td>
<td>Bx Prob</td>
<td>Alternating</td>
</tr>
<tr>
<td>Hallahan et al. (1982)</td>
<td>8</td>
<td>2</td>
<td>Glance</td>
<td>TR</td>
<td>Att Px</td>
<td>Continuous</td>
</tr>
<tr>
<td>Lloyd et al. (1982)</td>
<td>10</td>
<td>1</td>
<td>Glance</td>
<td>TR</td>
<td>Att Px</td>
<td>Alternating</td>
</tr>
</tbody>
</table>

Table 7 continued
<table>
<thead>
<tr>
<th>Study</th>
<th>Obs/Day</th>
<th>Number of Bx Coded</th>
<th>Number of Observers</th>
<th>Definition of Off-task</th>
<th>Diagnostic Method</th>
<th>Diagnostic Moniker</th>
<th>Observation Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rapport et al. (1982)</td>
<td>6</td>
<td>1</td>
<td>NR</td>
<td>Partial &gt;2s</td>
<td>SSI/T RS</td>
<td>ADDH</td>
<td>Continuous</td>
</tr>
<tr>
<td>Rhode et al. (1983)</td>
<td>5</td>
<td>15</td>
<td>NR</td>
<td>Whole</td>
<td>P&amp;T RS</td>
<td>Bx Prob</td>
<td>Continuous</td>
</tr>
<tr>
<td>Shepp &amp; Jensen (1983)</td>
<td>2</td>
<td>1</td>
<td>NR</td>
<td>Whole</td>
<td>T RS</td>
<td>Hyperactive</td>
<td>Continuous</td>
</tr>
<tr>
<td>Rooney et al. (1984)</td>
<td>5</td>
<td>1</td>
<td>NR</td>
<td>Glance</td>
<td>TR</td>
<td>Att Px</td>
<td>Alternating</td>
</tr>
<tr>
<td>Rosen et al. (1984)</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>Partial (any)</td>
<td>T RS</td>
<td>Hyperactive</td>
<td>Alternating</td>
</tr>
<tr>
<td>Rapport et al. (1985)</td>
<td>3</td>
<td>1</td>
<td>NR</td>
<td>Partial &gt;2s</td>
<td>P&amp;T RS</td>
<td>Hyperactive</td>
<td>Continuous</td>
</tr>
<tr>
<td>Burlton-Bennet et al. (1987)</td>
<td>10</td>
<td>4</td>
<td>2</td>
<td>Partial (any)</td>
<td>P&amp;T RS</td>
<td>Hyperkinetic</td>
<td>Continuous</td>
</tr>
<tr>
<td>Rapport et al. (1987)</td>
<td>3</td>
<td>1</td>
<td>NR</td>
<td>Partial &gt;2s</td>
<td>P&amp;T RS</td>
<td>ADDH</td>
<td>Continuous</td>
</tr>
<tr>
<td>Walker (1989) D</td>
<td>3</td>
<td>1</td>
<td>NR</td>
<td>Partial</td>
<td>T RS</td>
<td>ADHD</td>
<td>Continuous</td>
</tr>
<tr>
<td>Szymula (1990) D</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>Partial</td>
<td>P&amp;T RS</td>
<td>ADHD</td>
<td>Continuous</td>
</tr>
<tr>
<td>Bloomquist et al. (1991)</td>
<td>3</td>
<td>3</td>
<td>NR</td>
<td>NR</td>
<td>SSI/P&amp;T RS</td>
<td>ADHD</td>
<td>NR</td>
</tr>
<tr>
<td>DuPaul et al. (1992)</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>Partial (any)</td>
<td>P&amp;T RS</td>
<td>ADHD</td>
<td>Continuous</td>
</tr>
<tr>
<td>Paniagua (1992)</td>
<td>10</td>
<td>8</td>
<td>NR</td>
<td>Partial &gt;15s</td>
<td>T RS</td>
<td>ADHD</td>
<td>Continuous</td>
</tr>
<tr>
<td>DuPaul &amp; Henningson (1993)</td>
<td>6</td>
<td>2</td>
<td>NR</td>
<td>Partial &gt;3s</td>
<td>P&amp;T RS</td>
<td>ADHD</td>
<td>Continuous</td>
</tr>
<tr>
<td>Kraemer (1994) D</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>Partial</td>
<td>P&amp;T RS</td>
<td>Hyperactive</td>
<td>Alternating</td>
</tr>
<tr>
<td>Rapport et al. (1994)</td>
<td>3</td>
<td>1</td>
<td>NR</td>
<td>Partial &gt;2s</td>
<td>SSI/P&amp;T RS</td>
<td>ADDH</td>
<td>Continuous</td>
</tr>
<tr>
<td>Kelley &amp; McCain (1995)</td>
<td>4</td>
<td>3</td>
<td>NR</td>
<td>Partial (any)</td>
<td>P&amp;T RS</td>
<td>ADHD</td>
<td>Continuous</td>
</tr>
<tr>
<td>Turner (1996) D</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>Partial &gt;2s</td>
<td>SSI/T RS</td>
<td>ADHD</td>
<td>Continuous</td>
</tr>
<tr>
<td>Matheson (1997) D</td>
<td>3</td>
<td>2</td>
<td>NR</td>
<td>Partial</td>
<td>P&amp;T RS</td>
<td>ADHD</td>
<td>Continuous</td>
</tr>
<tr>
<td>Cloward (2000) D</td>
<td>5</td>
<td>1</td>
<td>NR</td>
<td>NR</td>
<td>T RS</td>
<td>ADHD</td>
<td>Continuous</td>
</tr>
<tr>
<td>Swenson et al. (2000)</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>Glance</td>
<td>NR</td>
<td>ADHD</td>
<td>Continuous</td>
</tr>
<tr>
<td>Roberts et al. (2001)</td>
<td>NR</td>
<td>1</td>
<td>NR</td>
<td>Partial</td>
<td>T RS</td>
<td>Bx Prob</td>
<td>Continuous</td>
</tr>
<tr>
<td>Northup &amp; Gulley (2001)</td>
<td>4</td>
<td>5</td>
<td>NR</td>
<td>Partial</td>
<td>NR</td>
<td>ADHD</td>
<td>Continuous</td>
</tr>
<tr>
<td>Austin (2003) D</td>
<td>15</td>
<td>10</td>
<td>NR</td>
<td>Partial</td>
<td>P&amp;T RS</td>
<td>ADHD</td>
<td>Continuous</td>
</tr>
<tr>
<td>Fabiano &amp; Pelham (2003)</td>
<td>22</td>
<td>2</td>
<td>NR</td>
<td>Whole</td>
<td>P&amp;T RS</td>
<td>ADHD</td>
<td>Continuous</td>
</tr>
<tr>
<td>Lorah (2003) D</td>
<td>7</td>
<td>5</td>
<td>2</td>
<td>Partial &gt;3s</td>
<td>SSI/P&amp;T RS</td>
<td>ADHD</td>
<td>Continuous</td>
</tr>
<tr>
<td>Dobrinski (2004) D</td>
<td>5</td>
<td>1</td>
<td>4</td>
<td>NR</td>
<td>T RS</td>
<td>Att Px</td>
<td>Alternating</td>
</tr>
<tr>
<td>Trolinder et al. (2004)</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>NR</td>
<td>T RS</td>
<td>Att Px</td>
<td>Continuous</td>
</tr>
<tr>
<td>Clarfield &amp; Stoner (2005)</td>
<td>2</td>
<td>3</td>
<td>NR</td>
<td>Whole</td>
<td>SSI/P&amp;T RS</td>
<td>ADHD</td>
<td>Continuous</td>
</tr>
</tbody>
</table>
Notes. ¹ Sample size reported as total number subjects whom met inclusion criteria for the current study and for whom data was reported; ² Mean and/or SD estimated from graph; A, Age; Bx Prob, Referred for behavioral problems/hyperactivity; C, Classroom; Continuous, Observing one child exclusively for entire observation period; Definition of Off-task, Proportion of interval needed to be coded off task; F, Females; G, Grade; Glance, Observers looked briefly at child and coded on/off-task for that instant; M, Multiple informants; No of Bx Coded, Number of behaviors coded; NR, Not reported; Obs/Day, Number of observation days; Observation Duration, Minutes of observation; Observation Interval, Seconds before recording; Observation Method (Alternating, abab or similar method; P, Parent rating scale Only; PR, Pediatrician referral; RS, Rating scale(s); S, Sex; SE, Special education classroom; SSI, Semi-structured interview; T, Teacher rating scale only; TR, Teacher Referral.
Figure 1: Original metric differences between observed classroom attention of children with ADHD and control children based on best case estimation after controlling for methodological differences among studies.
CHAPTER FOUR: CONCLUSION

Direct observations of children with ADHD and their peers in classroom settings indicate significant deficiencies in their ability to pay attention, and these differences are apparent regardless of most methodological differences across studies. Children with ADHD were also more variable in their attentive behavior, as indicated by significantly larger standard deviations across studies for ADHD compared to control groups. The current meta-analysis quantified and examined standardized effect size estimates of these differences, and found significant heterogeneity across studies that could not be explained by study-level sampling error. A weighted regression analysis of potential moderators determined that several factors, both methodological and subject-specific, have significant effects on the observed levels of attentive behavior in the classroom. Methodological differences, including sample characteristics, diagnostic issues, and observational coding schema, accounted for a large percentage of between-study variability such that no systematic variability remained across studies (adjusted R² = .91).

An overall mean weighted effect size of 0.71 was found before considering the impact of methodological variables. The significant heterogeneity among the 23 studies, however, suggests that this mean may not be the best indicator of actual differences between children with ADHD and control children. A best case estimation obtained by solving the significant moderator regression equation using values corresponding to best practice suggests that the actual standardized mean difference between the ability of children with and without ADHD to pay attention in the classroom is likely closer to 1.40. Original metric analysis of this finding suggests that across studies, typically developing children are off-task approximately 15% of observation periods compared to 38% for children with ADHD. Stated differently, children with ADHD are able to focus their attention in classroom settings approximately 62% of the time, compared with
an 85% on-task average for typically developing children. Power analysis based on Cohen’s (1992) recommendations suggests that only four subjects per group are needed to detect an effect of this magnitude.

Only two between-group studies failed to find significant differences in the classroom attentive behavior of children with ADHD and typically developing children. Cunningham and Siegel (1987), which approached significance at $p = .052$, utilized a design in which each ADHD child was paired with a typically developing child, and these dyads worked together for the duration of the study. The ADHD-Typically Developing child interaction may have contributed to the failure to find a significant difference in attention between the groups. The finding of no significant on-task differences in Shecket and Shecket (1976) may have been influenced by their decision to code only one of 19 possible behaviors during each observation interval. For example, a child coded as “Appropriate Talk with Teacher” would not have been coded as “Attending” during that same interval, potentially decreasing substantially the base rate of the attending category. It is unknown whether a re-analysis combining all codes consisting of on-task behavior would have resulted in significant between-group differences. No studies reported children with ADHD as being more attentive than control children.

Regression analysis revealed that sampling variables, diagnostic practices, and observational methods influence obtained results and must be considered when designing or implementing a direct observation system for children with ADHD and their peers. The impact of these moderators is exemplified by the best case approximation: Subject matching, more rigorous diagnostics, and longer observation of fewer behaviors were associated with an average effect size of 1.40 (Note: the mean across studies without controlling for methodology was 0.71).
Single case design (SCD) studies generally corroborate the between-group findings, although SCD studies were more likely than between-group studies to include children with ADHD that were off-task more than half of observed intervals. This difference does not appear to be clinically significant, however, and may be accounted for by several factors (Jacobson and Truax, 1991). Based on area under the curve analysis, the larger off-task rates in the small-\( n \) SCD studies can be predicted by the between-group distribution. The average between-group distribution predicts that approximately six subjects per study will score beyond the SCD mean, whereas on average only about 4 SCD subjects per study actually meet or exceed this value. No significant differences were found between effect sizes predicted by the between-group regression equation and obtained effect sizes for the six SCD studies with a typically developing comparison group, although obtained off-task rates were generally higher than predicted. The study types differed on three moderators found to significantly influence observed differences between the off-task rates of children with ADHD and their peers. Two of these moderators, number of behaviors coded and days of observations, were associated with larger effect sizes, whereas observation method was associated with smaller effect sizes. Selection bias may have also influenced the observed, though clinically insignificant, trend towards larger effect sizes for SCD studies. Single case designs, like case studies, often target children with the most extreme behavior problems for intervention. Publication bias may have also favored studies of children whom are more frequently off-task, as small-\( n \) designs have lower power and require larger effect sizes for statistical significance. No statistically significant difference in off-task rate, however, was found between published (\( M = 53.73, \ SD = 16.63 \)) and unpublished (\( M = 45.1, \ SD = 15.35 \)) SCD and between-group studies, \( t(18) = 1.71, p = .104, ns. \)
Matching subjects based on age, grade, and classroom at school were associated with relatively lower effect size estimates. Subject matching is designed to control for myriad extraneous variables including developmental level and the cognitive and situational demands placed on children in academic settings (Tabachnick & Fidell, 2001). The present findings corroborate this general trend and suggest that matching for relevant characteristics may be an important consideration when conducting classroom observation research. The larger effect sizes associated with natural classroom settings compared to simulation classrooms may be related to the myriad uncontrolled situational factors present in the former (Rapport, Timko, Kofler, Sims, & DuPaul, under review). Simulated classrooms tend to include highly trained (e.g., special education) teachers, multiple adult staff members, smaller class sizes (i.e., low student-teacher ratio), occur during irregular times of the year (i.e., summer), and frequently involve robust behavior management systems and highly structured daily routines (e.g., Pelham et al., 2000; Solanto et al., 2001).

Studies using increasingly more rigorous diagnostic criteria (for both number of informants/settings and addition of comprehensive diagnostic interviews) were associated with smaller overall effect sizes. This finding highlights the need for comprehensive diagnosis of research subjects and suggests that screening measures such as rating scale cut-off scores may be insufficient for assigning children to ADHD and control groups in research studies (Barkley et al., 1990). No behavioral cluster is pathognomonic to ADHD without considering onset, course, and duration of symptomatology, and systematically ruling out other possible causes or factors associated with impaired functioning. Structured clinical interviews provide improved sensitivity and specificity compared to rating scales, which typically assess a rater's perception of a limited, recent period of time (e.g., last 6 months; McClellan & Werry, 2000; Rutter & Graham, 1968).
Systematic differences in obtained effect sizes were also dependent on the diagnostic moniker assigned to the experimental group (i.e., hyperactive versus ADHD/ADDH). Earlier studies of hyperactive children tended to report moderately larger effect sizes compared to newer studies diagnosing children as ADHD/ADDH. This finding may reflect the change from a monothetic (DSM-III-R) to a polythetic (DSM-III; DSM-IV Combined Type) categorical diagnostic scheme, wherein children with other disorders were more likely to meet diagnostic criteria using the former system (August & Garfinkel, 1993). Children with ADHD may be more variable in their behavioral responses than children carrying other psychiatric diagnoses (e.g., Barkley, 1991). Samples containing subgroups of relatively more homogeneous groups of children – for example, children with anxiety or conduct problems – will have less variability and therefore larger effect sizes, as variability is the denominator in effect size formulae. More recent studies using gold standard assessment procedures and polythetic criteria may provide purer, and thus more variable, samples of children with ADHD. Inclusion of non-ADHD children may result in decreased group-level variability.

Observation code idiosyncrasies also influence observed differences between the ability of children with ADHD and typically developing controls to remain attentive in the classroom. Longer observation intervals and a greater number of observation days were associated with larger effect sizes, whereas the duration a child must remain off-task before (s)he is coded as such did not predict between-group differences. The high frequency of off-task behavior in children with ADHD may be responsible for the failure of the latter to incrementally influence effect size. These findings are consistent with some (Ary & Suen, 1983; Bramlett & Barnett, 1993), but not all (Mehm & Knutson, 1987; Smith, Madsen, & Cipani, 1981) observational research with children. None of the previous research, however, included samples of children with ADHD. Both
within-subject and between-group studies of ADHD and typically developing control children consistently reveal that children with ADHD are more variable in their ability to remain on-task in classroom and laboratory settings (Barkley, 1991). Longer observation intervals and a greater number of days are more likely to provide a valid, more reliable sampling of classroom attention. Scrutiny of within-subject experimental designs reveals that children with ADHD may exhibit 1-2 days of moderate to high on-task behavior followed by highly variable or low on-task days (e.g., Rapport et al., 1982). Heyman and colleagues (2001) provide a simple statistical method based on split-half reliability for determining the total observation duration necessary based on the frequency of behaviors of interest (see their Appendix, pp. 119-120). This method first calculates current reliability and then provides a formula to determine the number of minutes the observation needs to be increased/decreased to obtain a desired reliability. Adjusting the formula will also allow the experimenter to calculate the affect of varying observation intervals on overall reliability.

Studies alternating observations between ADHD and control children reported larger differences in the classroom attentional processes of ADHD and control children compared to studies observing one child per observation session. This difference may be related to the frequency of behavior observed. The present review suggests that children with ADHD are off-task approximately 38% of observation periods compared to 15% for control children. Available research indicates that observed differences may be dependent on incident frequency and coding scheme (Harris & Lahey, 1982; Haynes, 2001). Labeling a behavior as frequent or infrequent, however, is not based on accepted, standardized criteria. The attentive behavior of children with ADHD appears to be dissimilar from their peers in the classroom primarily during structured academic tasks (Porrino et al., 1983). Children with ADHD show a significantly different pattern
of attentive behavior over time during structured academic tasks compared to their peers, with the former showing more frequent shifts from on- to off-task and off- to on-task states (Rapport et al., 2005). Alternating between children within an observation session, although allowing for collection of temporally similar peer data, increases the probability of omitting these frequent shifts during coding (Harris & Lahey, 1982).

The number of behaviors coded simultaneously was linearly and positively related to the magnitude of attentional differences between ADHD and control children. Each behavior coded requires the coder to recall a specific, operational definition (e.g., Abikoff, 1977). Most (Jones, Reed, & Patterson, 1974; Mash & McElwee, 1974; Taplin & Reid, 1973), but not all (Frame, 1979) studies report an inverse relationship between the number of behaviors coded simultaneously and recording accuracy. The failure of Frame (1979) to find this relationship may be related to overall low agreement scores and lack of a criterion-referenced comparison (Harris & Lahey, 1982). The use of written or computerized coding sheets may help reduce these cognitive demands due to cuing and the ability to reference definitions (Abikoff et al., 1977). Increasing the number of behaviors coded also requires the observer to attend to different features of the child's behavior (e.g., watching the child's head/eyes, feet, and hands, listening to words/vocalizations). An interaction with the defined coding interval is also likely, with a higher number of behavioral codes requiring more observation time. For example, recording 13 behaviors during a 15-second interval may be more reliable than coding them every five seconds.

Several caveats are in order. The restricted age range of reviewed studies did not allow for examination of differences between younger and older elementary school children. Attention, cognition, and activity level show significant developmental trends based on extant literature (DuPaul & Stoner, 2004) and the current finding of systematic differences in observed classroom
attention when matching for age. We were also unable to determine the influence of specific academic task or differences during structured compared to unstructured classroom activities, because most studies either observed children across several tasks or did not report this information. Extant studies show significant variability in the on-task behavior of children with ADHD due to differences in classroom variables (e.g., Barkley et al., 1980; Flynn & Rapoport, 1976; Jacob et al., 1978; Whalen et al., 1978; Zentall, 1980), some of which may be associated with level of cognitive demand and/or working memory (Denney, Rapport, & Chung, 2005).

Girls with ADHD were underrepresented in available studies, which may have influenced the failure to find a relationship between gender and observed differences in classroom attention. Gender differences in behavioral and cognitive symptom prevalence are documented in most, but not all studies of children referred for ADHD, with girls tending to exhibit more inattentive and fewer hyperactive symptoms than their male counterparts (Abikoff et al., 2002; Biederman & Faraone, 2004; Carlson et al., 1986; Graetz et al., 2005; Seidman et al., 2005; Weiss et al., 2003; Yang et al., 2004). Interobserver agreement, a form of reliability, was not assessed because of discrepancies in reported metrics (e.g., percent agreement, phi coefficient, $r$ correlation, and kappa were reported in 6, 8, 5, and 3 between-group studies, respectively, with one study not reporting) and because all studies reported values above 0.8 for their metric.

The presence of comorbid diagnoses was not assessed in the current study. None of the reviewed studies included comorbid diagnoses beyond LD, ODD, or CD. The presence of comorbid disruptive behavior disorders has been found to predict increased symptoms of aggression, interference (e.g., interrupting class, talking out of turn), and inattention in the large MTA study (Abikoff et al., 2002). Comorbid anxiety disorders have been hypothesized to inhibit impulsive/hyperactive behaviors in children with ADHD (Quay, 1997), however, no effects on
any directly observed behaviors were found in the large-scale MTA study (Abikoff et al., 2002; Newcorn et al., 2001). Only Roberts (1990) included a psychiatric comparison group (aggressive children), limiting our ability to conclude that inattentive symptoms are pathognomonic to ADHD, and not merely a byproduct of general psychopathology.

We did not examine potential interaction effects among the moderator variables. For example, coding interval may interact with the number of behaviors coded, such that longer intervals mediate detrimental effects of attempting to simultaneously code multiple behaviors. Weighted regression analysis in SPSS 12.0, however, requires simultaneous entry of dependent variables, thus precluding traditional methods for testing interaction effects (Tabachnick & Fidell, 2001). It is statistically unlikely that the addition of interaction terms or other potential moderators would provide incremental benefit, considering the large amount of variance explained by the current model.

Results of the meta-analytic review indicate that direct observation studies of comprehensively diagnosed ADHD and control children in natural classroom settings that match subjects based on age and classroom, collect data over several days while limiting the number of target behaviors, and observe subjects consecutively for longer intervals, are likely to obtain results consistent with the best case estimation reported here. The coding schema, however, depends on the data needed. Researchers may wish to use more codes for investigational purposes, whereas clinicians seeking to supplement interview and rating scale data may be interested in only a few select behaviors. In the former case, it is recommended that researchers use videotaped observations, which allow multiple viewings and limit the number of behaviors that must be coded at one time. It is noted, however, that the between-group studies analyzed here
used live observations, therefore the incremental validity of videotaped sessions was not directly assessed.

Studies of children with ADHD reveal robust deficiencies in their ability to maintain attention in the classroom. Classroom observations, however, record only output – the child’s observable behavior – informing us only of the existence of an underlying problem, not its cause or nature. Problems sustaining attention may be a core deficit of ADHD, as conceptualized in the current DSM (APA, 2000), or they may be secondary to other, more primary deficits.

Deficits in behavioral inhibition may lead to problems with sustained attention (Barkley, 1997). An inability to suppress prepotent responses to stimuli has been argued repeatedly based on laboratory experiments such as the Stop Signal task. However, a recent meta-analysis found that such differences between children with ADHD and typically developing children could be accounted for by differences in primary reaction time, with children with ADHD reacting slower and more variably than their peers (Alderson, Rapport, Kofler, & Timko, under review).

Problems with one or more aspects of working memory may account for attentional deficits seen in children with ADHD (see Rapport, Chung, Shore, & Isaacs, 2001 for a theoretical overview). Preliminary evidence suggests robust deficiencies in the ability of children with ADHD to recall short lists of phonological and short groups of visual-spatial stimuli for more than a few seconds, compared to both typically developing peers and individuals with other psychiatric diagnoses (Rapport, Timko, Kofler, & Alderson, 2005). Inattention and motor activity increase linearly for all children as cognitive demands on working memory increase, although children with ADHD exhibit a disproportional increase compared to both typically developing children and children diagnosed with other psychiatric disorders (e.g., depression, anxiety). These rates of change are similar across phonological and visual-spatial working memory tasks (Rapport et al., 2005).
Classroom studies suggest that children with ADHD are significantly more hyperactive than their peers during in-seat classroom activities, but not during recess or other nonacademic periods (e.g., Porrino et al., 1984). It is theoretically possible that children’s movement serves a functional purpose such as maintaining or enhancing arousal to sustain attention and/or stimulate working memory. Inchoate neurophysiological evidence supports this theory – areas of the prefrontal cortex, but not the motor loop, are consistently implicated in fMRI studies of ADHD (Castellanos, 2001; Diamond, 2000). Abnormalities in motor loop (i.e., basal ganglia and associated areas) functioning would suggest that hyperactivity is ubiquitous in ADHD, however, findings of primarily prefrontal abnormalities point to problems with arousal, working memory, and the planning of behavior.

None of the reviewed studies examined the effect of time or attention by time interactions in children with ADHD, with or without a typically developing comparison group. The failure to examine time effects in classroom studies of attention severely limits our understanding concerning possible underlying mechanisms and processes responsible for the attentional difficulties associated with ADHD (e.g., whether they reflect a more general deficit in attenotional mechanisms or particular deficiencies in the ability to maintain attentional focus over time). Specialty statistical processes, including growth mixture modeling and wavelet analyses, will allow researchers to characterize both inter- and intra-individual changes in attention over time. These analyses from the signal processing literature allow researchers to group participants based on temporal patterns and analyze patterns in frequency and amplitude of behavior at the level of the individual (Castellanos et al., 2005).

The current meta-analysis provides unequivocal evidence for the existence of classroom attention deficits in children with ADHD. It also establishes the magnitude of these problems.
compared to their peers, providing estimates of the classroom attentional capabilities of both developmentally typical children and those diagnosed with ADHD. It is clear that children with ADHD exhibit marked deficiencies in their classroom attentional ability, and the current study provides substantial support for the inclusion of inattention as a core and pervasive symptom of ADHD in the DSM-IV (APA, 2000) model. Currently, no single rating scale or diagnostic tool possesses adequate positive and negative predictive power to accurately diagnose ADHD. Inchoate research suggests that variables including seatwork completion/accuracy and verbally intrusive behaviors may discriminate ADHD/non-ADHD children better than independent observations of on-task behavior (see Pelham et al., 2005 for a review). It is unknown, however, whether the use of the best case methods and standard differences in attention across all studies presented here will allow improved diagnostic accuracy and predictive power.
LIST OF REFERENCES

Studies preceded by an asterisk were included in the between-groups meta-analysis

Studies preceded by a ^ were included in the SCD analyses

*Abikoff, H., & Gittelman, R. (1984). Does behavior therapy normalize the classroom

*Abikoff, H., & Gittelman, R. (1985). The normalizing effects of methylphenidate on the


*Abikoff, H., Gittelman, R., & Klein, R. (1980). Classroom observation code for hyperactive

Observed classroom behavior of children with ADHD: Relationship to gender and

Forms & Profiles*. Burlington, VT: University of Vermont Research Center for Children,
Youth, & Families.

Alderson, R.M., Rapport, M.D., Kofler, M.J., & Timko, T.M. (under review). Attention-
Deficit/Hyperactivity Disorder and Behavioral Inhibition: A Meta-Analytic Review of the
Stop-Signal Paradigm.


