8-2001

# The Negative Halo Effect of Oppositional Defiant Behaviors on Teacher Ratings of ADHD: Impact of Child Gender

David A. Jackson

# THE NEGATIVE HALO EFFECT OF OPPOSITIONAL DEFIANT BEHAVIORS ON TEACHER RATINGS OF ADHD: IMPACT OF CHILD GENDER

by

David A. Jackson
Master of Arts, University of North Dakota, 1998

A Dissertation

Submitted to the Graduate Faculty

of the

University of North Dakota

in partial fulfillment of the requirements

for the degree of

Doctorate of Philosophy

Grand Forks, North Dakota
August
2001

This dissertation, submitted by David Jackson in partial fulfillment of the requirements for the Degree of Doctorate of Philosophy from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.

Alan R. King, Ph.D (Chairperson)

Myrna Olson, Ph.D.

Thomas Petros, Ph.D.

John D. Tyler, Ph.D.

Andrea Zevenbergen, Ph.D.

This dissertation meets the standards for appearance, conforms to the style and format requirements of the Graduate School of the University of North Dakota, and is hereby approved.

Dean of Graduate School

3-8-01

Date

## PERMISSION

| | |
|---|---|
| Title | The Negative Halo Effect of Oppositional Defiant Behaviors on Teacher Ratings of ADHD: Impact of Child Gender |
| Department | Psychology |
| Degree | Doctorate of Philosophy |

In presenting this dissertation in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my dissertation work or, in his absence, by the chairperson of the department or the dean of the Graduate School. It is understood that any copying or publication or other use of this dissertation or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my dissertation.

Signature 

Date 2/27/01

TABLE OF CONTENTS

CHAPTER

LIST OF TABLES

ACKNOWLEDGEMENTS

I am greatly indebted to a host of people who gave their time and talents to this project. First, I would like to thank the administration, staff and students of Grand Forks Christian School in Grand Forks, North Dakota who provided the child actors and classrooms necessary to make the tapes. Camera expertise was provided by Charles Crummy and the tapes were masterfully edited by David Waterman. Sharyl Simeone devoted hours casting the students and directing the scenes. Gary Jones, director of the northwestern Minnesota Area Special Education Cooperative, helped recruit teachers and provided the opportunity to utilize interactive television to reach other teachers in distant school districts. Much credit goes to Howard Abikoff, Ph.D. who created the original scripts and tapes and made them available for our use. Beth Gnagy, Ph.D. provided much needed technical assistance with the Classroom Observations of Conduct and Attention Deficit Disorder observation code.

I am also grateful to my advisor Alan R. King, Ph.D. who devoted so much of his time to me throughout this project, and to my other committee members, John D. Tyler, Ph.D., Andrea Zevenbergen, Ph.D., Thomas Petros, Ph.D., and Myrna Olson, Ph.D., each of whom contributed valuable editorial comments to this paper. Thanks also to the University of North Dakota for the financial support in the form of a summer dissertation fellowship.

Last, but not least, I would like to express a heartfelt "I love you" to my wife, Crystal L.K. Jackson, M.Ed. for her editorial and financial assistance, motivational support, and faith that I could complete this project.

ABSTRACT

This study explored one potential reason for differences in diagnostic rates of attention-deficit hyperactivity disorder (ADHD) between genders: teacher-based rating bias. Abikoff, Courtney and Pelham (1993) showed elementary teachers two tapes of a male child in a fourth grade classroom, then had the teachers complete ADHD and ODD rating scales. One tape depicted a normal child; another depicted a child exhibiting either ADHD or ODD behaviors. Rating comparisons from the ADHD v. ODD tapes showed biases: the ADHD tape was rated higher than the ODD tape on ADHD rating scales and lower on ODD rating scales; while the ODD tape was rated higher than the ADHD tape on ODD rating scales but equal on ADHD rating scales. It was hypothesized that ODD behaviors exerted a halo effect on ADHD ratings.

The present study replicated and extended Abikoff et. al's study with new tapes including female actresses, hypothesizing that bias existed with the male, but not the female tapes. Following the procedures of Abikoff et al., this study showed new tapes to 80, rural Midwestern teachers. Though the tapes followed Abikoff's scripts, objective behavioral rating scales found crucial differences between his tapes and the present study tapes.

ADHD v. ODD tape comparisons showed no bias. Yet, comparisons of ADHD/ODD vs. normal tape ratings showed a bidirectional bias: ADHD behaviors inflated ODD ratings, with females rated significantly higher on ODD behaviors than males, and ODD behaviors inflated ADHD ratings, with males rated significantly higher than females on ADHD behaviors.

Results indicate that teachers may not differentiate between ADHD and ODD behaviors on rating scales, and that gender of the child exhibiting disruptive behaviors influences teacher ratings. Since diagnosticians and prevalence rate studies rely upon teacher ratings, these findings imply: (a) compared to females, the male prevalence rate for ADHD may be artificially inflated by the presence of ODD behaviors; (b) compared to males, the female prevalence rate for ODD may be artificially inflated by the presence ADHD behaviors; and (c) the comorbidity rate between ADHD and ODD may be artificially inflated by teacher failure to differentiate between ADHD and ODD behaviors.

# CHAPTER 1

## INTRODUCTION

Attention-deficit hyperactivity disorder (ADHD) might be the most commonly diagnosed childhood, school-related behavior problem in the United States (Wolraich & Baumgaertel, 1997). In fact, according to Sabatino and Vance (1994), this disorder has reached epidemic proportions in some regions of the country. Yet, despite extensive research (Barkley, 1998) relatively little is known about this often controversial condition (Wolraich & Baumgaertel, 1997). The nosology of the disorder has been in a state of constant flux since its inception, reflecting changing ideas about etiology and the advancing technology of diagnostic procedures (Garfinkel & Amrami, 1992).

Descriptions of children fitting this condition have been traced back as far as 1848 when German physician Heinrich Hoffman wrote a book for his children in which he described "Fidgety Phil", and "Harry Look in the Air", names suggesting hyperactivity and inattentiveness, respectively (Hoffman, 1948; as cited in Wolraich & Baumgaertel, 1997). Presenting a series of three lectures to the Royal College of Physicians in England in 1902, clinician George Still described 43 defiant, aggressive, overactive, and highly-emotional children who also exhibited significant problems with sustained attention. Still (1902) attributed these behavioral problems to a "defect in moral control" and contended that the defect resulted from one of three types of impairments: "(1) defect of cognitive relation to the environment; (2) defect of moralconsciousness; or (3) defect in inhibitory volition" (p. 1011).

In North America, interest in ADHD-type behaviors is often traced to an encephalitis epidemic in 1917-1918. During this period, physicians were faced with children who survived encephalitis, yet were left with cognitive and behavioral problems such as impaired attention, impulsivity and overactivity. These children, who clearly had suffered from brain damage, were often diagnosed with "Postencephalitic Behavior Disorder" (Barkley, 1998). In 1937, Bradley discovered that such behaviors could be managed with amphetamines.

ADHD has long been associated with academic and other school-related difficulties; as a result, the study of the disorder in the United States is closely linked to the study of learning disabled children. Prior to the 1940's, learning difficulties were thought to result from one of three conditions: mental retardation, emotional disturbances, or social and cultural disadvantages (Silver, 1992). In the early 1940's, a fourth cause was identified: a nervous system disorder. Researchers identified children who looked normal, yet presented with learning problems similar to children with brain damage. Accordingly, it was thought that such children suffered from minor brain damage, and the term, "minimal brain damage", was coined. During the 1940's the concept of the "brain-injured child" proliferated to the point where it was assumed that any psychiatrically hospitalized child with ADHD-type symptoms suffered from brain damage, whether or not there was a documented history of brain pathology (Barkley, 1998). However, this conceptualization was short-lived.

Due to a paucity of research supporting the minimal brain damage categorization, a competing theory arose during the 1950's. This second theory contended that the learning and behavioral problems exhibited by these

children resulted from "faulty wiring" within the brain. That is, all of the brain mechanisms appeared present and operable, but some of the nerve pathways were not functioning correctly. For example, Laufer, Denhoff and Solomons (1957) contended that children with ADHD-type symptoms suffered from a dysfunction in the thalmic region of the central nervous system. The idea of brain dysfunction, as opposed to brain damage, eventually became the accepted view, and the term "minimal brain dysfunction" (MBD) was coined (Silver, 1992).

In 1963, the National Society for Crippled Children and Adults, in collaboration with the Neurological and Sensory Diseases Service Program of the Division of Chronic Diseases of the U.S., formed a task force to review MBD. In the resulting document, children with MBD were described as follows:

> Children of near-average, average, or above average general intelligence with certain learning or behavioral disabilities ranging from mild to severe, which are associated with deviations of function of the central nervous system. These deviations may manifest themselves by various combinations of impairment in perception, conceptualization, language, memory, and control of attention, impulse, or motor function (Clements, 1966, p. 1114).

The committee also discussed the emotional and social problems associated with MBD. Using modern terminology to translate the 1966 findings, Silver (1992) stated that the committee had described children with MBD as having: (a) learning disabilities; (b) hyperactivity, distractibility, impulsivity; and (c) emotional and social problems. In the original language, the committee had delineated at least 99 symptoms for MBD. In essence, they had defined a

"vague and over-inclusive" category of "little or no prescriptive value" (Kirk, 1963, as cited in Barkley, 1998). Over time, the MBD concept faded away, as it became evident that these stimulant-responsive children formed a heterogeneous group, who shared no gross neurological deficits, but did share various degrees of distractibility, inattentiveness, clumsiness, impulsivity, aggressiveness, and learning difficulties (Weinberg & Brumback, 1992).

The first edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM; APA, 1957) did not address this disorder, but DSM-II (APA, 1968) introduced Hyperkinetic Reaction of Childhood. In this conceptualization, hyperactivity, or excessive body movement, was the primary deficit associated with the disorder. The DSM-III (APA, 1980) shifted focus away from hyperactivity and towards a primacy of attention deficits. The condition was renamed attention deficit disorder (ADD) and the manual distinguished between two subtypes: ADD with hyperactivity (ADD/H) and ADD without hyperactivity (ADD/WO). The DSM-III took a monothetic diagnostic approach. The manual recognized three behavioral dimensions: inattention, impulsivity and hyperactivity. The diagnosis of ADD/H required three symptoms of inattention, three symptoms of impulsivity, and two symptoms of hyperactivity. This system resulted in a more restrictive set of criteria in which fewer children were identified. In addition, those identified tended to be more severely affected (Garfinkel & Amrami, 1992).

The DSM-III-R (APA, 1987) changed the name to attention-deficit hyperactivity disorder (ADHD) and added the category of undifferentiated attention-deficit disorder (UADD) to replace ADD/WO. The condition was considered unidimensional and the manual used a polythetic system requiring

the presence of any eight of 14 items as the threshold for the diagnosis. In addition, by making ADHD a unidimensional construct, DSM-III-R avoided trying to categorize each symptom under a unique domain, and was able to remain consistent with the criteria for other DSM-III-R disorders. However, removal of the three behavioral dimensions decreased specificity and resulted in a more inclusive, heterogeneous diagnosis (Morgan, Hynd, Riccio, & Hall, 1996).

Desgranges, Desgranges, and Karsky (1995) stated that the DSM-III-R gave little guidance to prevent the confusion of other disorders with attention-deficit hyperactivity disorder. The DSM-IV (APA, 1994) addressed these types of concerns by adding three clarifying criteria: (a) symptoms must be present in two or more situations (e.g., at school, work, and home); (b) disturbance causes clinically significant distress or impairment in social, academic, or occupational functioning; and (c) disorder does not occur exclusively during the course of a pervasive developmental disorder, schizophrenia, or other psychotic disorder, and is not better accounted for by a mood disorder, dissociative disorder, or a personality disorder. In addition, based on factor analytic studies examining how the different behaviors grouped together (Garfinkel & Amrami, 1992), the DSM-IV defined two dimensions: inattention, and hyperactivity/impulsivity. From these two dimensions, three subtypes of ADHD were conceptualized: ADHD, predominately inattentive type; ADHD, predominately hyperactive/impulsive type; and ADHD, combined type. The new diagnostic subtypes of the DSM-IV have increased reported prevalence rates of the condition (Baumgaertel & Wolraich, 1995; Lahey et al., 1994) while decreasing the heterogeneity present in the DSM-III-R's definition.

The DSM-IV's inattention category includes the following nine criteria: (a) makes careless mistakes, (b) has difficulty sustaining attention, (c) seems not to listen, (d) fails to finish tasks, (e) has organization difficulties, (f) avoids tasks requiring sustained attention, (g) loses things, (h) becomes easily distracted, and (i) experiences forgetfulness. The hyperactivity category lists six criteria: (a) fidgeting, (b) difficulty remaining seated, (c) moving excessively, (d) difficulty engaging in leisure activities quietly, (e) being "on the go", and (f) talking excessively. The impulsivity category includes three criteria: (a) blurting answers before questions are completed, (b) difficulty awaiting turn, and (c) interrupting/intruding upon others (APA, 1994).

## Assessment Considerations

A multi-method assessment approach is recommended for the diagnosis of and treatment planning for ADHD (Barkley, 1998; Schaughency & Rothlind, 1991). Barkley (1998) suggested that physicians keep several goals in mind when evaluating children for the disorder. The first objective should be to determine the presence or absence of ADHD, including the differential diagnosis of ADHD from other childhood psychiatric disorders such as anxiety, depression, and oppositional defiant disorder. A second goal is to begin formulating treatment approaches to address the academic, psychological and social impairments resulting from the disorder. A third objective is to determine the presence of comorbid disorders and how they may contribute to the child's problems and impact the prognosis for treatment. The final goal is to obtain a pattern of the child's psychological strengths and weaknesses and formulate how these may impact treatment.

Ideally, the multi-method assessment should include a medical examination and clinical interview of the child. A diagnostic interview with the parents, completion of behavior ratings scales by the parents and a teacher, direct observation of the child's behavior, and the administration of clinic-based tests should also be included (Barkley, 1998). Numerous methods have been developed to try to measure the constructs of attention and hyperactivity/ impulsivity. Of these approaches, behavior checklists and rating scales have been most widely used due to their low cost, ease of administration and the wealth of information which may be obtained from them. Such scales include the Conners' Teacher Rating Scale (CTRS; Conners, 1969), the Child Behavior Checklist (CBCL; Achenbach, 1978), the Behavior Problem Checklist (Quay & Peterson, 1983), the ADD-H Comprehensive Teacher's Rating Scale (ACTeRS; Ullmann, Sleator, & Sprague, 1985), the IOWA Conners (Loney & Milich, 1982), the Swanson, Nolan and Pelham Rating Scale (SNAP; Atkins, Pelham, & Licht, 1985) the Disruptive Behavior Disorder Rating Scale (DBD; Pelham, Gnagy, Greenslade, & Milich, 1992), and the Child Attention Profile (Barkley, 1990).

In addition to rating scales, objective, direct observational procedures have also been developed, such as the Stony Brook Observation Code (Kent & O'Leary, 1976) and the Classroom Observations of Conduct and Attention Deficit Disorders (COCADD; Atkins, Pelham, & Licht, 1985). According to Barkley and Edelbrock (1987), direct observation procedures provide objective data less swayed by biasing factors, such as halo effects, that tend to influence parent and teacher reports. Standardized, clinic-based measures of sustained attention and impulse control have also become common in the assessment of ADHD. The first measure of sustained attention and vigilance, the Continuous

Performance Test (CPT; Rosvold, Mirsky, Sarason, Bransome, & Beck, 1956), provided the general format for numerous other similar tests. The Matching Familiar Figures Test (MFFT; Kagan, 1966) is currently one of the most widely employed clinic-based measures of impulsivity (Barkley, 1998).

## The Barkley Critique

Despite improvements in psychology's diagnostic armamentarium, the search for a valid and reliable measure of ADHD has not yielded an acceptable "litmus test" or "gold standard" (Barkley, 1998). Many practicing physicians, school teachers, parents, and the laity at large mistakenly believe that a positive response to methylphenidate or other psycho stimulants provides such a test, demonstrating the validity of the diagnosis (Weinberg & Brumbach, 1992; Golden, 1992). On the contrary, stimulants produce the same results in normal children and adults as they do in individuals with ADHD (Swanson, McBurnett, & Wigal, 1993; Wolraich & Baumgaertel, 1997). In the absence of diagnostic clarity, clinicians are implored to proceed with caution, using an exhaustive multi-method assessment approach. Barkley (1997) has even suggested that an adequate assessment of ADHD could take up to a year.

Barkley (1998) has hypothesized three major reasons why ADHD is so difficult to diagnose. First, the core symptoms of the disorder lie at the heart of human nature: we are all prone to episodes of inattention, impulsivity and overactivity. This creates a problem as the symptoms themselves are present, to some degree, in nearly everyone. Some researchers have contended that even the three behavioral domains themselves are elusive, defy strict definition and are too broad to convey any information of real value (Sunder, 1992). For example, Mesulam (1985, as cited in Sunder, 1992, p. 455) defined attention as

"the climax of mental integration", while Sherrington (1940, as cited in Sunder, 1992, p. 455) suggested that attention is "the most important prerequisite for the manifestation of intellectual and reflective powers." In essence, attention constitutes a readily apparent component of all human activity and behavior. Viewed in this light, Sunder (1992) stated that the term attention-deficit has no specificity and is a gross oversimplification: "It is an attempt to reduce multiple cognitive processes and effects to a single disordered output, an oversimplification of complex higher cortical activity, a 'reductio ad absurdum'" (p. 455). Sunder, like numerous other clinicians and researchers, believes that ADHD is a heterogeneous group of disorders, similar to the epilepsies, and that the label elevates a group of descriptive symptoms to a diagnosis about as precisely as a the label "headache" elevates a throbbing head into a medical entity.

The prevailing assumption has been that an aggregate of traits or symptoms is somehow pathognomonic for ADHD. Due to the lack of a valid and reliable laboratory test for the diagnosis, we must rely on behavioral criteria and checklists regarding behaviors that are present in nearly everyone at some time. Levine (1992) pointed out that this checklist approach presents at least three serious inferential flaws: First, there is the assumption that the greater number of symptoms present, the more likely one is to have the disorder. Second, questions arise regarding where the lines of demarcation should be drawn between the absence and presence of the disorder. And third, the observations themselves are subjective. Unlike hallucinations in schizophrenia, the three behavioral domains of ADHD have not been shown to be pathognomonic for the disorder.

Barkley's (1998) second contention as to why ADHD is so difficult to assess is that certain ADHD symptoms are typical of the full range of psychiatric disorders. For example, inattention is found in numerous disorders and, if anything, represents a global marker for distress. Measures of inattention have generally failed to distinguish children with ADHD from other psychiatrically impaired children (Halperin, Matier, Bedi, Sharma, & Newcorn, 1992; Shapiro & Garfinkel, 1986; Werry, Elkind, & Reeves, 1987). Likewise, measures of impulsivity, such as the Matching Familiar Figures test (Kagan, 1966) and the Porteus Maze test (Porteus, 1965), have also not consistently distinguished between ADHD and other patient groups (Koriath, Gualtieri, Van Bourgondin, Quade, & Werry, 1985; Loney & Milich, 1982; Tant & Douglas, 1982).

The third reason Barkely (1998) presented for the assessment difficulties is that ADHD is not an "all-or-nothing" condition like schizophrenia, cancer and pregnancy, but rather, Barkley views ADHD as being at the extreme end of the normal curve of human behavior. Consequently, the assessment of ADHD is ultimately a subjective endeavor in which psychologists and psychiatrists determine what degree of deviance represents a disturbance and then solicit teachers' and parents' opinions regarding the degree to which the child is disturbed. Golden (1992) stated that herein lies a critical assumption made by society: if a behavior is disturbing then it must be disturbed. Golden contended that society then attempts to control the behavior, or make it more palatable, by using a medical diagnosis, which then mandates certain pharmacological treatments. Golden connoted that it appears more acceptable to give a child in this broad group the "medical" diagnosis of ADHD and provide pharmacological intervention, than it is to accept the behavior for what it is: disruptive. Szasz

(1974, 1992) has spent the past 25 years writing about the medicalization of "problems in living" which result in behaviors that disturb others.

If the prevalence of ADHD is increasing, as some authors suggest (Garfinkel & Amrami, 1992; Ingresoll, 1988; Lapouse & Monk, 1958; Silver, 1993) than it would appear that the threshold between disturbed and undisturbed behavior has been lowered. The Drug Enforcement Administration (DEA) production quotas for methylphenidate (Ritalin) in the United States increased from 1768 kg in 1990 to 10,410 kg in mid-1995 (Safer, Zito, & Fine, 1996). Since 90% of methylphenidate is prescribed for ADHD, this increase suggests a possible six-fold increase in individuals receiving medication for ADHD, which would suggest a large increase in the diagnosis of ADHD. According to Safer and Zito (1999), an estimated 2.5 million youths in the United States take stimulants, primarily for the treatment of ADHD. Desgranges, Desgranges, and Karksky (1995) suggested that our increasingly sedentary and fatigued society has come to view normal childhood activity as excessive. On the other hand, Block (1977) and Ross and Ross (1982) believe that the increase in ADHD may be the result of an increasingly rapid "cultural tempo", reflected in the increased rates of stimulation and change in Western culture. Barkley (1998) pointed out that the belief that ADHD has increased is "difficult to address because no community surveys of ADHD have been repeated in the same populations or geographic areas over sufficiently long periods to evaluate for such trends" (p. 85).

## Methodological Considerations

Social critics (Kohn, 1989; Schrag & Divoky, 1975; Weinberg & Brumbach, 1972) have long argued that professionals are quick to label

vivacious and exuberant children as being mentally disordered. These critics have also charged that educators may use labels such as ADHD as an excuse for poor educational environments, as an excuse for a general inability to deal with difficult children, and a reason to obtain additional funds. Desgranges et al. (1995) noted that the minute a child begins misbehaving in the classroom, someone labels the problem ADHD. The authors suggested that the symptom list from the DSM have become popularized to the point where people feel that anyone exhibiting the symptoms must have ADHD. The general public does not realize that the diagnosis should only be made if the symptoms are present and other causes are not able to explain the situation (Desgranges et al.).

A disorder which cannot be strictly defined nor precisely and objectively measured is difficult to study. Adding the facts that the ADHD diagnosis is behaviorally driven, that the behaviors of interest are present to some extent in nearly everyone, that the criteria overlap with those of other psychiatric conditions, that the line of demarcation between normal and abnormal is subjectively drawn, and that those reporting on the behaviors may have something to gain through the rendering of a diagnosis, makes the study of ADHD extremely difficult. Clearly, given this state of affairs, the prevalence of the condition cannot be accurately measured (Barkley, 1998). Yet, the consensus of expert opinion seems to be that approximately 3 - 5% of the childhood population has ADHD (American Psychiatric Association, 1994). However, estimates from epidemiological studies have ranged from 1 - 20% (Dupaul, 1991; McGee, Williams, & Silva, 1987; Ross & Ross, 1982; Rutter, Tizzard, & Whitmore, 1970; Szatmari, Offord, & Boyle, 1989).

There appear to be at least five important variations in measurement across these studies which help account for the prevalence differences (Szatmari et al.,1989) First, not all studies have used the same symptoms to define the disorder. Most studies have used inattention and overactivity as defining criteria. Other studies, however, have used symptoms such as low frustration tolerance, temper tantrums, irritability, and negative peer interactions to define the construct (Miller, Palkes, & Stewart, 1973; Nichols & Chen, 1980; Werner et al., 1968), apparently reflecting the changing conceptualization of ADHD over the decades.

Second, studies have used varying methods of data collection. Rutter et al. (1970) used parent and teacher checklists as well as psychiatric interviews. Nichols and Chen (1980) used behavior ratings made by psychologists during testing. However, most studies have used only teacher rating scales (Pelham, Nagy, Grenslade, & Milich, 1992; Szatmari et al., 1989; Trites, Dugas, Lynch, & Ferguson,1979; Wolraich, Hannah, Pinnock, Baumgaertel, & Brown, 1996).

Third, the sources of information have differed across studies. Studies have used parents (e.g., Skekim et al., 1985; Werner et al., 1968; ), teachers (e.g., Pelham et al., 1992; Schachar, Rutter, & Smith, 1981; Szatmari et al., 1989; Trites et al., 1979; Wolraich et al., 1996), physicians (Lambert, Sandoval, & Sassone, 1978) and children (Shekim et al., 1985) both alone and in various combinations as informants. Lambert et al. reported rates varying from 1% to 13% depending upon the source of information.

Fourth, the criteria or threshold score to make a diagnosis has varied in restrictiveness, resulting in the identification of different numbers and types of cases. For example, Trites et al. (1979) and Nichols and Chen (1980) only

required scores in excess of 1 and 1.5 standard deviations above the mean, respectively, on teacher-reported measures to make the diagnosis of ADHD. In contrast to these broad definitions, Rutter et al. (1970), Glow (1981), and McGee, et al. (1985) employed tighter restrictions, stipulating that children with both hyperactivity and a comorbid disorder would not be considered hyperactive for the purpose of their studies.

Finally, differences in sample characteristics across studies have led to major differences in prevalence rates. For example, Rutter et al.'s (1970) Isle of Wight study was undertaken on a rural population of 10-11 year old children. In contrast, the Nichols and Chen study (1980) was on a birth cohort of 7-year-olds from an urban area. In addition, most studies (Glow, 1980; Lambert et al., 1978; Miller et al., 1973; Skekim et al., 1985; Szatmari et al., 1989; Trites et al., 1975) have been done on public school samples, neglecting home and private school children.

Although prevalence studies have used varying symptoms to define the disorder, varying methods to collect the data, varying sources from which to obtain information, varying criteria and thresholds for a diagnosis, varying samples with varying characteristics to study, and have suggested varying prevalence rates, one finding has been nearly universal: males with ADHD outnumber females with ADHD. Some researchers have concluded that ADHD is primarily a male diagnosis (Gaub & Carlson, 1997). Indeed, the earliest description of children suspected of suffering from ADHD was of males: "Fidgety Phil" and "Harry Look in the Air" (Hoffman, 1948; as cited in Wolraich & Baumgaertel, 1997). Prevalence rates vary, with male to female ratios in clinical samples ranging from 9:1 to 6:1 (APA, 1994). Ratios from population-

based samples are consistently lower at approximately 3:1 (APA; Szatmari, Offord, & Boyle, 1989) and have been found to be as low as 2.1:1 (Taylor, Hepinstall, Sonuga-Burk, & Sandberg,1998).

Greater male prevalence also appears in conduct disorder (CD), while the third of the DSM-IV's disruptive behavior disorders of childhood, oppositional defiant disorder (ODD), seems to present an age dependent pattern. Anderson, Williams, McGee and Silva (1987), studying 11 year-olds, found males with ODD to outnumber females with ODD at the rate of 2.2:1. Likewise, Cohen, Valez, Kohn, Schwab-Stone and Johnson (1987) found a 2.3:1 male preponderance in their sample of 9-12 year-olds. However, studying adolescents, Kashani et al. (1987) found females to outnumber males 2:1; while McGee et al. (1990) found a 3:1 female predominance.

## Comorbidity Data

Data from both clinical and epidemiological populations have suggested that comorbidity among the disruptive behavior disorders (DBD) is high. Hinshaw (1987) estimated the overlap between hyperactivity and aggression to be between 30% and 90%. Although lower estimates of comorbidity have been found in non-referred samples, such studies have still revealed a large overlap. Recently, August, Realmuto, MacDonald, Nugent, and Crosby (1996) screened 7,321 school children for the presence of DBD's and found that of the children with ADHD, 32% presented with ODD and 12% with CD. The few studies that have evaluated sex-differential comorbidity, although inconsistent, have suggested less externalizing pathology (e.g., antisocial, aggressive and oppositional behaviors) in females with ADHD as compared to males with ADHD (Berry, Shaywitz, & Shaywitz, 1985; Eme, 1992; Gaub & Carlson, 1997);

however, this trend was not found in the DSM-IV field trials (Lahey, 1994). Other studies (e.g., Conners, 1994, as cited in Arnold, 1996) have found more internalizing (anxiety, depression) comorbidity in females with ADHD as compared to males with ADHD.

## Gender Differences in ADHD

"Why then are more boys identified as hyperactive/ADD than girls" (McGee, Williams, & Silva, 1987, p. 711)? Berry et al. (1985) stated, "Reasons for this gender discrepancy remain obscure." (p. 801). Two logical conclusions may be drawn regarding the large gender-based prevalence difference in ADHD: either the prevalence differential reflects the true base rate of the disorder, or the difference is an artifact of some type of bias. If the former condition is true, then it should be reasonable to propose that either the etiological pathways to the disorder are different across males and females or the pathways are similar but there are gender differences in the prevelance of the pathway conditions (e.g., differences in hormones or neurotransmitter levels). Differences in pathways or pathway conditions should be reflected in a different pattern of correlates of ADHD across males and females (McGee & Feehan, 1991). If no clear differences appear, then the latter condition must be seriously considered. However, consistent differences found between males and females with ADHD must be viewed through the lens of naturally occurring gender differences in non-afflicted individuals. The genders normally differ on several measures relevant to ADHD; therefore, differences between ADHD males and females must be different in kind, or exceed these naturally ocurring differences, before they are considered as differences in the expression of the disorder.

Parent and teacher rating scales have routinely demonstrated that girls of all ages are less hyperactive and have fewer attention problems than same-aged boys, and such findings have been cross-culturally consistent (Achenbach, 1991; Bauermeister, 1992; Brito, Pinto, & Lins, 1995; Conners, 1994, as cited in Arnold, 1996; Goyette, Conners & Ulrich, 1978; Trites, Blouin, & Laprade, 1980). On both structured diagnostic interviews and rating scales, normal females have shown fewer symptoms of both ADHD and other externalizing problems, including delinquency and aggression, than have their normal male counterparts (Achenbach, 1991; Bauermeister, 1992; Conners, 1994, Eme, 1992; Garb & Carlson, 1997; Lahey, 1994; Zoccolillo, 1993). On neuropsychological measures, such as the Continuous Performance Test, normal girls have also been shown to make fewer errors, have slower reaction times and faster, but no less accurate, digit cancellation scores than boys (Pascualvaca, 1994, as cited in Arnold, 1996). Brain imaging studies have also shown females to have smaller brains, but larger caudate nuclei than males (Giedd et al., 1996).

## Cognitive Functioning

Several studies have found that females with ADHD perform poorer on cognitive tasks and have lower IQ scores than their male counterparts (Berry, Shaywitz , & Shaywitz, 1985; Brown, Madan-Swain, & Baldwin, 1991; Gaub & Carlson,1997; James & Taylor, 1990; Kashani, Chapel, Ellis & Shekim, 1979). Numerous other studies, however, have not found statistically reliable differences on cognitive measures and IQ scores between males and females with ADHD (Arcia & Connors, 1998; Breen, 1989; Horn, Wagner, & Lalongo, 1989; McGee, Williams, & Silva,1987; Sharp et al., 1999). Barkley (1989)

suggested that the lower cognitive functioning found in ADHD females, relative to ADHD males, might result from the sources from which ADHD research participants have been obtained: "Where children were drawn from pediatric learning problem clinics, greater cognitive impairment was found in girls. . . However, where subjects were chosen from a clinic specializing in hyperactive children, few sex differences were noted" (Barkley, p. 380). Similarly, in their paper describing their meta-analysis, Gaub and Carlson (1997) hypothesized that the poorer intellectual abilities found in ADHD females, relative to ADHD males, might be "restricted to clinic-referred children" (p.1041). The authors noted that no studies had compared male and female IQ scores using non-referred samples of ADHD subjects. Gaub and Carlson also proposed four additional methodological problems with the literature which confounded their ability to draw solid conclusions regarding gender differences among individuals with ADHD. By logical extension, these same problems would also apply to any attempts to draw conclusions regarding gender-based cognitive and intellectual differences among ADHD-afflicted individuals.

The first of the four problems was that differential comorbidity patterns between the genders could influence intellectual functioning. Of the 18 studies reviewed by the Gaub and Carlson (1997) only nine addressed comorbidity, and most those studies assessed for only a limited number of potential comorbid disorders. Second, most of the studies inadequately addressed developmental considerations: only 2 of the 18 studies reviewed in the meta-analysis evaluated for the effect of the children's ages. Third, inadequate and inconsistent diagnostic procedures often resulted in comparisons between dissimilar subjects. Fourth, the literature rarely allowed for the adequate

evaluation of the effects of rater source on the examination of gender differences and ADHD. With these methodological problems in mind, the paper will review some of the studies which have compared males and females with ADHD on various intellectual and cognitive measures.

Learning Disorder Clinics and Hospital-Based Referrals

Kashani, Chapel, Ellis and Shekim (1979) screened 740 children at a pediatric developmental evaluation clinic over a three-year period. Of these children, 298 were diagnosed with hyperkinetic reaction of childhood. After additional exclusionary criteria were applied, 28 hyperkinetic girls remained in the study. These girls were then matched with hyperkinetic boys on socioeconomic class, race and age. The 56 children in the sample were evaluated on a variety of measures, including the Wechsler Intelligence Scale for Children-Revised (WISC-R; Wechsler, 1974). The results of the study indicated that that hyperkinetic girls scored significantly lower on the Verbal Scale, compared to hyperkinetic boys. In addition, evaluations by a pediatric neurologist revealed significantly more language disabilities in the hyperkinetic females compared to their male counterparts. The reasons for referral to the evaluation clinic were also significantly different between the genders, with boys more frequently referred for hyperactivity and behavioral disorders, while most girls were referred for learning disabilities, and language and speech disorders.

Berry et al. (1985) compared 32 girls with attention deficit disorder (ADD) to 102 similarly-diagnosed boys. Subjects were recruited from a university-based learning disorders unit and a pediatric neurology clinic. Each child received a comprehensive evaluation, including a neurologic examination.

Intelligence quotient scores from the WISC-R were used when they were available in the children's' school records, which resulted in IQ comparisons between 20 girls and 79 boys. The results indicated that the girls scored significantly lower on the Verbal Scale from the WISC-R, had significantly poorer ratings on academic and language abilities, and experienced significantly higher referral rates for speech problems, compared to their male counterparts. The authors suggested that cognitive deficits should play a more prominent role in the identification of girls with ADD.

Breen (1989) evaluated 13 girls and 13 boys recruited from referrals to a pediatric psychology clinic. The children, who were all diagnosed with ADHD, were administered the Gordon Diagnostic System (GDS; Gordon, 1983), an instrument purported to measure sustained attention and impulse control, as well as the subtests from the Kaufman Assessment Battery for Children (K-ABC; Kaufman & Kaufman, 1983), thought to be the most sensitive to academic readiness, memory and attention. The study found no significant differences between the genders on any of the nine scales administered.

James and Taylor (1990) retrospectively studied the charts of 61 males and 18 females who had attended the Maudsley and Bethlem Royal Hospitals between 1968 and 1982. All of the children had received the ICD-9 diagnosis of hyperkinetic syndrome of childhood. James and Taylor found that girls with hyperkinetic syndrome had lower IQ's, as measured by the WISC-R, the Wechsler Preschool and Primary Scale of Intelligence (WPPSI; Wechsler, 1967), and the Merrill-Palmer (Ball, Merrifield, & Scott, 1978), as well as significantly higher rates of language and other neurological disorders, compared to the boys with the syndrome. The researchers proposed that the

ICD-9 diagnostic category of hyperkinesis was more often associated with the presence of an "organic brain disorder" in girls than in boys with hyperkinesis. Noting that significantly more of the females than males had been seen by pediatricians prior to their hospital referral, the authors suggested that the females may have been displaying more "neuropsychiatric problems."

ADHD Clinic-Based Referrals

Brown, Madan-Swain, and Baldwin (1991) reported data from 51 ADHD boys and 20 ADHD girls treated at their university based ADHD clinic. The authors evaluated the primary symptoms of ADHD: inattention, hyperactivity, and distractibility, as well as the secondary symptoms of learning problems and academic difficulties. Brown et al. reported significant gender differences on only 2 of 36 measures: compared to boys, the girls scored lower on the Block Design subtest from the WISC-R and lower on the Spatial Memory subtest of the Simultaneous Processing scale of the K-ABC. In addition, the researchers found that girls had been retained in school more frequently, and were nearly one year older than their male counterparts at the time of treatment referral.

Arcia and Connors (1998) analyzed WISC-R and WAIS-III test results from 132 males and 27 females who had been seen at a private, university-based ADHD clinic. The subjects, who had all been diagnosed with ADHD, ranged between the ages of 5 - 60. The results indicated that the male and female IQ scores showed no statistical differences.

Horn, Wagner and Lalongo (1989) compared 37 ADHD male and 17 ADHD female elementary school-aged children who had been referred to a university-based psychology clinic for treatment of chronic inattention and impulsivity. The authors administered the Peabody Picture Vocabulary Test-

Revised (PPVT-R; Dunn & Dunn, 1981) the Wide Range Achievement Test-Revised (WRAT-R; Jastak & Wildinson, 1984) and the Personality Inventory for Children -Revised (PIC-R; Wirt, Lachar, Klinedinst, & Seat, 1984). The results revealed no significant gender differences on the three WRAT-R subscales nor on the four PIC-R subscales that measure learning problems. [However, the males scored significantly higher than the females on the PPVT-R (which provides an IQ measure that correlates satisfactorily with the Full Scale from the WISC-R)].

Community and Non-Specific Referrals

McGee, Williams, and Silva (1987) evaluated the IQ scores of children involved in the Dunedin Multidisciplinary Health and Development Study, a longitudinal investigation of a sample of children born between April 1, 1972, and March 31, 1973, in Queen Mary Hospital. McGee et al's study selected 20 boys who received a score of 12 or higher on Rutter Child Scales A and B (Rutter, Tizzard, & Whitmore, 1970), and 17 girls who received scores of 7 or higher on the same measure. Different cutoff scores were used for boys and girls in order to identify children with attention problems by reference to the same-sex distribution of scores. Utilizing different cutoff scores resulted in the identification of 4.9% of the boys and 5% of the girls as being inattentive. Analysis of WISC scores revealed that, relative to their attentive peers, the inattentive boys and girls showed the same pattern of IQ deficits. However, compared to each other, the inattentive boys and girls scored nearly identical on the WISC.

Sharp et al. (1999) compared 42 girls with DSM-III-R / DSM-IV ADHD (combined type) to 56 similarly-diagnosed boys. The children had been

referred to the study from a variety of sources, including: schools (35%), physicians (24%), research program alumni (10%), the National Institutes of Health listing of clinical studies (10%), and friends (7%). The results revealed a slight trend for the girls scoring lower on the Verbal, Performance, and Full Scale Indices from the WISC-R; however, the gender differences were not statistically significant. In addition, the findings were somewhat confounded by the higher rates of comorbid disorders in the female subjects: oppositional defiant disorder (girls 50%, boys 33%, $p = .09$), major depression (girls 7%, boys 0%, $p = .08$), and specific phobia (girls 7%, boys 0%, $p < .01$).

Meta-Analysis and Conclusion

A meta-analysis by Gaub and Carlson (1997) reviewed 18 studies published between 1979 and 1992. In order to qualify for the analysis, each study had to meet five criteria: (a) direct comparison of boys with ADHD and girls with ADHD on relevant variables; (b) at least 10 subjects per group; (c) subjects had to be less than 14 years old; (d) subjects had to have IQ scores greater than 79; and (e) the study had to document adequately ADHD status using DSM-II, DSM-III, or DSM-III-R criteria. Following the meta-analysis, the authors concluded that clinic-referred girls with ADHD scored lower on Verbal, Performance and Full Scale sections of intelligence tests compared to clinic-referred boys. The analysis, however, utilized only 6 of the 18 studies to derive the Full Scale difference figure, and even fewer studies (3) to calculate the Verbal and Performance figures. In addition, the effect sizes for the Full Scale and Performance differences were small, while the effect size for the Verbal difference were moderate (Cohen, 1977).

Conflicting findings and methodological problems, such as those reviewed above, led Rhee, Waldman, Hay and Levy (1999) to suggest that no conclusions may be drawn regarding gender differences in cognitive impairment between boys and girls with ADHD. Likewise, there appears to have been a general disagreement over this issue at the National Institute of Mental Health's 1994 conference which had convened to discuss gender differences in ADHD (Arnold, 1996; Barkley, 1995).

## Symptom Externalization

The few studies which have examined sex-differential comorbidity for children with ADHD have been inconsistent (Arnold, 1996). Some studies have suggested less externalizing pathology in girls than boys. Berry, Shaywitz, and Shaywitz (1985), studying 32 girls and 102 boys with ADHD, concluded: "Inappropriate and antisocial behaviors characterized both girls and boys, but loss of control and physical aggression were more troublesome features of disorder in boys" (p. 806). However Berry et al. also stated that management problems and antisocial behaviors were found to be correlates of hyperactivity "irregardless of gender" (p. 808). Using 12 studies to calculate an effect size, Gaub and Carlson (1997) found that males with ADHD exhibited significantly higher rates of externalizing behaviors compared to girls with ADHD. In addition, the effect size calculated from ten studies suggested that non-referred females with ADHD suffer from more internalizing conditions than non-referred males with ADHD. The meta-analysis also showed that boys with ADHD, compared to their female counterparts, exhibited significantly greater levels of hyperactivity, inattention and peer aggression. Based on the results of Gaub and Carlson's meta-analysis and other data presented at the NIMH Conference

in 1994, Barkley (1995) concluded that there are, unequivocally, gender differences in comorbid conditions in children with ADHD. Specifically, males with ADHD exhibit more oppositional defiant and conduct disorder symptoms than females with ADHD, regardless of whether the sample is drawn from community or clinical populations.

However, findings stated above and supported by Barkley (1995) were not supported in the DSM-IV field trial data (Lahey et al., 1994). Recently, Sharp et al. (1999) compared 42 girls with DSM-III-R / DSM-IV ADHD to 56 comparably diagnosed boys. Parents rated attention problems as being more severe in girls while teachers rated boys as being more hyperactive; yet, no significant differences were found between the two groups for either internalizing or externalizing symptoms. Arcia and Conners (1998) analyzed clinic data on 360 non-medicated ADHD patients (280 male and 80 female) ranging in age from 5 to 60 years. The males and females did not differ significantly on parent or teacher ratings of hyperactivity, inattention, conduct disorders or internalizing behaviors. However, the self-ratings of adults did differ with females rating themselves as having significantly fewer personal "assets," such as the ability to make friends or the possession of a "sharp" mind, compared to their male counterparts. Arcia and Connors also found that boys and girls were similar in comorbidity, whether defined in general (ADHD plus at least one other diagnosis) or by individual analyses for all diagnoses that were present in either group. In a recent study of gender differences that employed the largest sample of girls to date, Biederman (1997) compared 130 girls with ADHD with 120 normal control girls. In terms of risk for comorbid DSM-IV psychiatric disorders, the girls with ADHD showed elevated rates of major

depression, anxiety and bipolar disorders, and these rates were not significantly different from rates the investigators had found using male subjects in a previous study. The only rates which differed from their previous studies with boys were the rates of oppositional defiant and conduct disorders in girls, which were approximately half the rates found in boys.

## Course and Prognosis

Research has suggested that there may be some gender differences in the life course of ADHD. Klein (1994; as cited in Barkely, 1995) reported that females in her longitudinal study were significantly younger at the time of initial referral, were more educated by the time they reached adulthood, were less likely to have been substance abusers or conduct disordered, and were less likely to remain diagnosed with ADHD at adulthood than the males in the study. Overall, Klein's findings suggested a better outcome for females with ADHD than for males. The generalizability of these findings, however, has been challenged based on the small sample size (i.e.,19 subjects). In contrast to Klein's findings, Arcia and Connors (1998) found no such difference in age of initial referral, and their adult female subjects reported significantly fewer "assets" and more functional problems with concentration, restlessness, anger, confidence, and feelings than males. The evaluation of adult clinical samples has led some to suggest that the large, gender-related prevalence difference dissipates later in life (Arnold, 1996). According to the Scott-Levin physician diagnosis and drug audit data, a higher proportion of ADHD office visits after age 17 are by females, though the referral rate falls for both sexes after age 17 (Williams & Swanson, 1994; as cited in Arnold, 1996).

## Treatment Responsiveness

In terms of treatments responsiveness, there has been a paucity of controlled medication trials on girls with ADHD. Barkley (1989) and Pelham, Walker, Sturges, and Hoza (1989) found no differences between boys and girls in response to methylphenidate. In the largest placebo-controlled comparison of methylphenidate and dextroamphetamine in girls with ADHD, Sharp et al. (1999) found that both girls and boys exhibited robust beneficial responses to both stimulants, with nearly 95% of the subjects responding favorably to one or both drugs. No studies have evaluated whether there are gender differences in response to psycho social treatments (Barkely, 1995); however, Arnold (1996) suggested that there may be reason to suspect differential effects. For example, it has been shown that mothers tend to be more critical of their ADHD daughters than their ADHD sons (Barkley, 1998) and this may have implications for parent training, e.g., techniques such as positive attending and "catching them being good" may need more emphasis in families in which the child with ADHD is female.

## Models of Etiology

Biologically-based etiological theories of ADHD have been proposed to account for the gender-based prevalence rate differential. Postulated causes for the differential have included the following: (a) greater male vulnerability to perinatal injuries (Ounsted, 1972), (b) relative male immaturity (Rutter, 1970), (c) different cerebral lateralization (Geshwing & Galaburda,1985), (d) male fetal antigenicity (Gaultieri & Hicks, 1985), (e) polygenetic inheritance with a higher threshold for expression in females (Cloniger, Christiansen, Reich, &

Gottesman, 1978), and (f) greater male constitutional variability (Taylor & Ounsted, 1972).

Regarding ADHD in general, studies have shown the disorder to have a large heritability component. Risk for ADHD is higher in the first-degree relatives of children with ADHD, compared to the non-first- degree relatives of children with ADHD. In addition, this risk is even greater for relatives of children with ADHD and antisocial behaviors, compared to the relatives of children with ADHD and no antisocial behaviors (Faraone, Biederman, Keenan, & Tsuang, 1991). Adoption studies have shown that the biological parents of children with ADHD are more likely to have ADHD than are the adoptive parents (Morrison & Stewart, 1973). There is a paucity of studies, however, examining sex differences in the magnitude of genetic and environmental influences on ADHD (Rhee, Waldman, Hay, & Levy, 1999). Goodman and Stevenson (1989) found similar heritabilities for boys and girls on both hyperactivity and inattention; Thapar, Hervas, and McGuffin (1995) reached the same conclusions. In a recent large scale study, Rhee et al. examined 2391 twin and sibling pairs from Australia. Once again, the researchers found the magnitude of familial and environmental influences on ADHD symptoms to be similar for both boys and girls.

Gualtieri and Hicks (1985) concluded that males are more frequently afflicted with neurodevelopmental disorders than are females; however, when the disorders arise in females, a more severe form is usually manifest. Two theories were devised to account for this seemingly paradoxical situation: the polygenetic multiple-threshold model (DeFries, 1989) and the constitutional variability model (Eme, 1992; James & Taylor, 1990; Taylor and Ounsted,

1972). Both models have been researched as explanatory agents for the gender-prevalence differential in ADHD. Thus, the paper will present both models, focusing on the predictions made by each model, as well as the research supporting and refuting each theory.

Polygenetic Multiple-Threshold Model

The polygenetic multiple-threshold model is based on the general multi-factorial model of disease transmission (Carter, 1969). Diseases are thought to be caused by numerous genetic and environmental factors and no assumptions are made about the relative magnitude of each potential source. All of these sources combine additively to determine one's liability or vulnerability for the disorder. Individuals in the population all have some degree of liability, and this variation in liability determines the population's liability distribution. The difference between one afflicted with a disorder and one not afflicted is that the former crossed the threshold of liability necessary for the manifestation of the disorder, while the latter's liability did not surpass the critical threshold. For disorders such as ADHD in which there are gender-based prevalence differentials, it is assumed that males and females have congruent underlying liability distributions, but females have a higher threshold than males.

Eme (1992) suggested that three predictions flow from the polygenetic multiple-threshold model. First, since females have higher thresholds, they must require a higher liability than males to develop the disorder. A corollary of this prediction is that female probands with the disorder should have more deviant manifestations of ADHD than male probands. This is because the higher liability needed to affect the females should result in more severe manifestations of the disorder (Tsai & Beisler, 1983). Second, females with

ADHD should have larger genetic loadings then males, as this would be needed to cross their presumably higher thresholds. Third, since females require greater liability to develop the disorder, those afflicted should transmit their greater liability to their offspring. Because of the greater transmitted liability, relatives of female probands with ADHD should carry a higher liability than relatives of affected male probands; consequently, there should be a higher prevalence of ADHD in relatives of female probands. In summary, this model suggest a continuum of liability with higher thresholds for females, who are predicted to be more severely affected, have higher genetic loadings, and to have more afflicted relatives than males.

Studies testing the polygenetic multiple-threshold model have resulted in conflicting conclusions. Kashani, Chapel, Ellis, and Shekim (1979), Pauls, Shaywitz, Kramer, Shaywitz, and Cohen (1983) and Faraone et al. (1995) found evidence in support of the model, while Mannuzza and Gittleman (1984), Goodman and Stevenson (1989), James and Taylor (1990), Silverthorn, Frick, Kuper, and Ott (1996) and Rhee, Waldman, Hay, and Levy (1999) found evidence against it. As an example of this type of research, Pauls et al. compared the risk of ADD in siblings of male and female probands. They found the risk to siblings for ADD was . 35 if the proband was female and .23 if the proband was male, providing evidence in support of the model. Goodman and Stevenson compared the pairwise concordances in hyperactive boys and girls and their dizygotic twin brothers and sisters. The boys and girls had nearly the exact same probability of being hyperactive if they had a hyperactive brother or a hyperactive sister, providing evidence against the model. Eme (1992)

contended, "the data necessary to adequately evaluate the polygenetic multiple-threshold model are virtually nonexistent." (p. 361).

One of the most persistent themes in male-female differentiation is their rate of development. From the middle of the fetal period onward, the average boy is behind the average girl in physical maturity (Taylor, 1985). This difference in maturation is 1 year at the beginning of school, 1 1/2 years at age 9, and 2 years at the onset of puberty (Tanner, 1978). According to Eme (1992) there are two major consequences of this slower rate of development in boys. First, males are more susceptible to pathology because immature organisms are at greater risk than mature organisms (Gualtieri & Hicks, 1985). In addition, because of the slower transcription of genomic information, there is a greater likelihood for variations in that transcription, and some of this mistranscription may cause pathology. Second, it is hypothesized that for females to develop the disorder, they must experience some type of pathological event, such as brain damage. Consequently, the affected female would experience greater divergence from her "norm" and thus, be more severely affected than a male.

Constitutional Variability Model

The constitutional variability model hypothesizes that relatives of male probands with ADHD should have higher prevalence rates for the disorder than the relatives of female probands. This is because female affliction is thought to be caused by a rare, non-genetic, organic "accident" while male affliction is thought to be genetic, and thus, more likely to occur. This is in direct contrast to the polygenetic multiple-threshold model which predicts that relatives of female probands should have a higher probability of ADHD.

Essentially, research supporting the constitutional variability model is comprised either of studies finding higher rates of ADHD for the relatives of male probands than the relatives of female probands, or studies showing greater cognitive impairments amongst ADHD females. Mannuzza and Gittleman (1984), Goodman and Stevenson (1989), James and Taylor (1990), Silverthorn, Frick, Kuper, and Ott (1996), and Rhee, Waldman, Hay, and Levy (1999) found evidence supporting this model, while Kashani, Chapel, Ellis, and Shekim (1979), Pauls, Shaywitz, Kramer, Shaywitz, and Cohen (1983) and Faraone et al. (1995) found evidence contradicting the model. James and Taylor (1990) found that girls diagnosed with ICD-9 hyperkinetic syndrome of childhood had lower IQ's and significantly higher rates of language and other neurological disorders than similarly diagnosed boys. They also found that females were characterized by high rates of medical illness affecting the brain. In addition, there were greater distributions of IQ's amongst the affected boys, with some boys faring quite well on standard batteries. The authors interpreted these findings as supportive of the constitutional variability model. Berry, Shaywitz and Shaywitz (1985), Kashani, Chapel, Ellis and Shekim (1979) as well as Gaub and Carlson's meta-analysis (1997) similarly found girls with ADHD to be more cognitively impaired than their male counterparts. Rhee, Waldman, Levy, and Hay (1999) pointed out that it is also possible to interpret these lower abilities as being supportive of the polygenetic multiple-threshold model. Essentially, girls with the disorder may have lower IQ's because they carry a higher biological loading of liability, which may cause not only ADHD, but other cognitive impairments as well. In addition, numerous studies have found no differences in cognitive functioning between males and females with

ADHD (Arcia, & Connors, 1998; Breen, 1989; Horn, Wagner, & Lalongo, 1989; McGee, Williams, & Silva,1987; Sharp et al., 1999).

In summary, a review of the literature regarding biological explanations for, and major correlates of, the gender-based prevalence differential in ADHD reveals inconsistencies and contradictions. There is no clear evidence of a qualitatively different expression of the disorder in males than in females (Barkley, 1995) and neither the polygenetic multiple-threshold nor constitutional variability models have been accepted as explanatory for the large prevalence rate difference. The conceptualization of ADHD has changed over the years, shifting focus from brain dysfunction to hyperactivity to inattention. The number and type of diagnostic criteria have changed and both monothetic and polythetic diagnostic systems have been employed. Assessment procedures have changed and various treatment fads have come and gone. Yet among these changes, one thing has remained constant: males diagnosed with ADHD have always outnumbered females diagnosed with ADHD. Without an acceptable biological explanation, and lacking consistent gender-based correlates, one must consider the possibility that the prevalence differential may be an artifact of gender bias.

## Gender Bias

According to Hartung and Widiger (1998), one of the most divisive issues in mental health diagnostics has been gender prevalence rates. Accurate estimates of gender prevalence are hard to obtain and are subject to dispute and controversy. Many of the conclusions regarding gender prevalence rates are complicated by common sources of bias. Widiger and Spitzer (1991) define bias as being " a systematic deviation from an expected value" and sex bias as

"a systematic deviation that is associated with the sex of the subject" (p. 2). Garb (1997) noted that a bias only occurs "when the accuracy of judgments varies as a function of the client's race, social class or gender" (p. 99). Widiger and Spitzer pointed out that differential prevalence rates resulting from biological differences should not be considered biased, as biological differences are inherent, reflecting actual, expected sex differences. Hartung and Widiger suggested that most of the error in prevalence studies may be attributed to either diagnostic sex bias or sampling sex bias.

A diagnostic sex bias occurs when there is a differential prevalence in either false positives or false negatives between the genders (Widiger and Spitzer, 1991). A false positive is the misdiagnosis of the presence of a disorder, while a false negative is the misdiagnosis of the absence of a disorder. According to Widiiger and Spitzer, two types of diagnostic sex bias may be identified. "Criterion sex bias" relates to the diagnostic criteria encoded in the official nomenclature of the DSM, while "assessment sex bias" addresses the assessment instruments and methods which provide the diagnoses. These include self-and other-report inventories, psychological test batteries, as well as clinical judgment.

## Gender-Based Expectations

A prime source of assessment sex bias is the clinician's own gender-based expectations, knowledge and values (Widiger & Spitzer, 1991). Loring and Powel (1988) concluded, "allegedly objective evaluations, even when guided by an intricate set of seemingly clear-cut criteria, can be influenced by characteristics of the observer making the judgments and of the individual being evaluated" (p. 17). Piel-Cook, Warnke, and Dupuy (1993) define gender bias

in diagnosis as occurring whenever expectations, values, or ignorance about gender result in inaccurate or incomplete diagnoses of clients. Male and female clients with the same symptomology have been shown to earn different diagnoses (Garb, 1997; Hamilton, Rothbart, & Dawes, 1986). Hamilton et al. had 65 licensed clinical psychologists independently diagnose 18 written case histories on the basis of ten DSM-III categories including antisocial and histrionic personality disorders. The results showed that females were rated significantly more histrionic than males exhibiting identical histrionic symptoms. There was no comparable sex bias towards diagnosing males showing antisocial pathology as more antisocial then females.

Analogue studies such as this - testing for differential diagnosis based on sex - are generally considered to be the least likely to yield evidence of bias or attributional error (Loring & Powell, 1986). After reviewing gender-related analogue studies, Abramowitz and Dokecki (1977) concluded, "Clinical analogues that have afforded a more direct test of the notion of evaluative prejudice against women have for the most part refuted it" (p. 63), while Sherman (1980) characterized the bulk of the studies as "consistent with bias and sex role stereotyping" (p. 51). Zeldow (1978, as cited in Loring & Powell, 1988, p. 6) summarized the literature by stating, "studies are sufficiently diverse and ambiguous as to be interpretable both as strong and weak evidence for sexism in the mental health field, depending on the point of view of the interpreter."

The second type of diagnostic gender bias, criterion sex bias, involves the DSM criteria. Diagnostic criteria are said to be biased if clinicians adhere to the criteria when making diagnoses, and still one gender receives a

disproportionate number of diagnoses. Hartung and Widiger (1998) stated; "Ideally, diagnostic criteria sets would be gender neutral. However, many of the criteria sets may disproportionately favor the manner in which the disorder appears in one gender relative to the other" (p. 267). Diagnostic neutrality is hard to obtain for disorders which express themselves differently between the genders, or involve maladaptive variants of gender-related behaviors. Gender bias due to diagnostic criteria has been well researched (Brown, 1992; Caplan, 1995; Walker, 1994). After studying 2,013 adolescents, Huselid and Cooper (1994) concluded the following, "Our findings replicated gender differences in patterns of symptom expression, with female adolescents reporting more psychological distress and lower self-esteem and male adolescents reporting more delinquency and substance use. More important, these results reveal that gender roles account for a substantial portion of the sex differences in both internalizing and externalizing symptoms" (p. 600).

Sampling Bias

According to Hartung and Widiger (1998) the second major source of error in prevalence studies involves sampling procedures. Most studies in mental health use non-probability samples of convenience which are highly susceptible to selection bias. Consequently, disproportionate representation of the sexes is common within clinical research (Gannon, Luchetta, Rhodes, Pardie, & Segrist, 1992) and may distort prevalence rates, ultimately creating problems in the development of unbiased diagnostic criteria. In terms of sampling bias, biased representation in both clinical settings and empirical studies needs to be addressed.

Within clinical settings, differential sex prevalence rates reflect many gender differences independent of the treated disorders. Such differences include the willingness to acknowledge problems and seek treatment, the reactions of others to one's symptoms, and the presence of comorbid conditions. The DSM-IV (APA, 1994) suggests factors such as these as explanatory for the large gender-based prevalence differences found between clinical and community samples. Hartung and Widiger (1998) point out that the DSM-IV fails to address the differential sex prevalence rate in disorders of childhood versus adulthood. The authors suggest that this difference reflects "differences in the behaviors or symptoms that are recognized as being disordered in childhood versus adulthood" (p. 263). Of the 21 disorders usually first diagnosed in infancy, childhood, or adolescence, 17 are said to be more common in boys, while only three are more common in girls. In contrast, of the 80 disorders of adulthood for which sex ratios are provided, 35 are said to be more common in men than women (17 of which are substance related or paraphilia) while 31 are said to be more common in women, and 14 have equal distributions. From this data it would appear that mental disorders are overwhelmingly more common in boys than girls, then the distribution evens out in adulthood.

This shift from male dominance in childhood clinical samples to a more even adult distribution has been attributed largely to the source of referral for treatment. In childhood, the motivation for treatment comes mainly from parents and teachers (Popper & Steingard, 1994). Consequently, the behaviors which lead to childhood clinic referrals tend to be those that are of concern to others (Shaywitz, Shaywitz, Fletcher, & Escobar, 1990). In contrast, adults tend to self-

refer, with perhaps more females seeking treatment than males (Good & Wood, 1995). In essence, the DSM-IV childhood disorders are more sensitive to symptoms which are troubling to others, while the adult sections are more sensitive to symptoms which are troublesome to the identified patient. Keenan and Shaw (1997) recently reviewed gender differences in the prevalence rate in some forms of childhood psychopathology. The authors concluded that prior to age seven and the beginning of school there are few gender differences in the rates of difficult temperament, activity level or noncompliance. This pattern shifts towards male predominance in elementary school, then towards female predominance during adolescence; by adulthood, females are twice as likely to seek treatment for the internalizing disorders of anxiety and depression compared to males.

Referral Bias

Referral bias due to comorbid conditions and disruptive/difficult behaviors heavily influences prevalence rates. Shaywitz et al. (1990) suggested that the predominance of boys with reading disorders may be an artifact of comorbid behavioral problems. They found that systematic structured assessments revealed few sex differences in the rate of reading disorders. Yet boys, due to disruptive behavior problems, were more frequently referred for services than girls. Likewise, Gaub and Carlson's meta-analysis (1997) comparing boys and girls with ADHD concluded that boys are more likely to be referred for services than girls due to the presence of comorbid externalizing disorders. The researchers also cautioned that clinic-referred girls with ADHD may not be representative of the general population of girls with ADHD. Barkley (1995) summarized the results from the National Institute of Mental

Health's conference which had convened to discuss gender differences in ADHD by stating, "A referral bias operates in determining which gender is likely to get referred for clinical services. This implies that females may require a more severe degree of ADHD and associated disruptive behaviors...than do males before such a referral is made" (p. 4). Werry and Quay (1971) concluded that boys are more "at risk" than girls to develop behavioral disorders, based on their findings that elementary school boys more frequently demonstrated symptoms of inattention, hyperactivity, and acting-out behaviors associated with "badness" than girls. Berry, Shaywitz and Shaywitz (1985) suggested that a selective referral bias represents an important factor in determining the predominance of males with ADHD. Specifically, boys with inattention, hyperactivity, and acting out behaviors are more likely than girls with these problems to be referred for services because their behaviors are "troublesome to adults" (p. 801). The typical referral comes from school once the boy has become a management problem in the classroom.

Since clinical settings do not represent an accurate reflection of the prevalence of a disorder, empirical studies drawing random samples from within a clinic will not be accurate either. As a result, clinic-based empirical research on disorders such as ADHD has been done nearly exclusively on boys. Goodman and Kohlsdorf (1994) cautioned that confining research to only one gender can contribute to an inaccurate (e.g., male-biased) description of the disorder. This, in turn, will impact the diagnostic criteria for the disorder, which will lead to even greater disproportionate sampling, which will further hone the diagnostic criteria to favor one sex (Robbins, 1991). Hartung and Widiger evaluated empirical studies in the Journal of Abnormal Child

Psychology from 1987-1994, the six-year period between publications of the DSM-III-R and the DSM-IV. Of 254 published studies, 11 concerned adults or did not specify the gender of the participants. Of the remaining 243, 71% (173) included both boys and girls, while 29% (70) were confined to just one sex. Of the 70 studies confined to one sex, 69 studied boys while only one studied girls. Seventy of the 243 empirical studies involved ADHD and only half of those studied both sexes, while 100% of the 36 unisex studies were confined to males. A total of 4,837 children participated in these 70 ADHD studies. Eighty-one percent of them were boys (3,967) and only 19% (906) were girls. Given statistics such as these, one may see how ADHD may have developed into a male-dominated disorder.

The discrepancy between community and clinical samples (APA, 1994), the finding that prevalence rates level in adulthood (Arcia & Conners, 1998; Arnold, 1996), and the conclusion that etiology and associated characteristics may be the same in both genders (McGee, Williams & Silva,1985) have led some to suggest that the gender-based prevalence difference in ADHD may be an artifact of bias (Arcia & Conners, 1998; Barkley, 1995; Berry, Shaywitz & Shaywitz, 1985; Gaub & Carlson, 1997; Henker & Whalen, 1989; McGee & Feehan, 1991; McGee, Williams, & Silva, 1987). Historically, the gender bias controversy has typically concerned false positive diagnoses in females (Widiger & Spitzer, 1991). In the case of ADHD, however, the controversy has involved just the opposite (i.e., false negative diagnoses of females).

Berry, Shaywitz and Shaywitz (1985) studied a group of 32 girls and 102 boys diagnosed with either pure attention deficit disorder (ADD) or attention deficit disorder with hyperactivity (ADD-H). Within the ADD-H group, girls had

more severe cognitive impairments, were younger at time of referral and came from families of lower socioeconomic status; disruptive, uncontrollable behaviors were more frequent in boys from this group. In addition, girls both with and without hyperactivity were more likely to suffer peer rejection than their male counterparts. These led the authors to conclude, "Girls with ADD may represent an under-identified and under-served group of children that is at significant risk for long-term academic, social, and emotional difficulties" (p. 808).

McGee, Williams and Silva (1987) selected inattentive boys and girls by reference to the same sex distributions of teacher ratings; that is, one group was comprised of the highest rated girls, while a second group consisted of the highest rated boys. Relative to their normal peers, inattentive boys and girls showed the same patterns of deficits on numerous cognitive measures as well as similar histories of behavioral problems at school. Had they not used same sex distributions for subject selection, most of the girls would not have been identified by the teacher ratings. The authors echoed the sentiments of Berry, Shaywitz and Shaywitz (1985) that girls are under-identified and under-served.

McGee and Feehan (1991) carefully reviewed data from three large epidemiological studies: the Isle of Wight Study (Rutter, Tizzard, & Whitmore,1970); the Dunedin Multidisciplinary Health and Development Study (McGee, Williams. & Silva, 1987) and the Ontario Child Health Study (Szatmari, Offord, & Boyle, 1989). The authors concluded that there were generally no differences between the genders in correlates of ADHD, etiological pathways, or treatment responsiveness, nor did girls require a "larger dose" of risk factors to develop the disorder. The authors found that sex differences in behaviors

were associated primarily with teacher ratings only. Furthermore, they recommended that it would be more appropriate to gauge ADHD in girls against the standard of what is normal behavior for girls in general, as opposed to comparing them to boys. In essence, they recommended the use of same-sex distributions when determining cutoff scores on measures of ADHD.

Gaub and Carlson's (1997) meta-analysis concluded that girls are less likely to be referred to clinics, and that those who are referred are the most severely affected. They based this assertion on the findings that girls with ADHD are less impaired then boys with ADHD in the general population, but equally impaired in clinical populations. The authors discussed the potential use of different cutoff scores for males and females. Barkely (1995) stated that the use of gender-based cutoff scores in making diagnoses was also discussed at the National Institute of Mental Health Conference on gender differences in ADHD. He concluded that there was widespread controversy on this issue.

## Prevalence Considerations

Bias refers to any deviation from an "expected value" (Widiger & Spitzer, 1991). In prevalence studies, the expected value is the actual base rate of the disorder in the population under study. Widiger and Spitzer discuss some of the problems inherent in determining the presence of a sex bias:

> Because bias involves a deviation from an expected value, determination of bias requires a knowledge of the expected value. This can be complicated, since the expected value for one form of sex bias will at times be the differential sex prevalence that results from another form of sex bias. For example, a sampling sex bias involves a deviation from a representative sampling of the population, but the differential sex

> prevalence that would be obtained by a representative sampling might have resulted in part from social-cultural etiological sex biases (e.g., more females than males have the disorder because more females than males are subjected to a particular etiologic trauma). Similarly, a diagnostic sex bias involves a deviation from the actual prevalence of the disorder in a particular setting, but this prevalence could reflect a sampling sex bias (e.g., more females than males with the disorder appear at the clinic). The differential sex prevalence that would be obtained by an unbiased diagnosis could then itself still be sex biased with respect to sampling (p. 3).

Though it is unlikely that the prevalence rate for any mental disorder will be identical across the genders (Meehl, 1967), in the absence of evidence to expect otherwise, gender prevalence rates should not be widely divergent. Much of this current paper has been directed toward presenting evidence suggesting that gender base rates for ADHD are not as disparate as prevalence studies have suggested. Historically, researchers have posited that females with ADHD have been under-identified and under-served. Possibly resulting from this line of research, the DSM-IV criteria for ADHD were broadened substantially resulting in the diagnosis of previously unidentified youth at the rate of 4:1, with the newly identified cases twice as likely to be female (Lahey et al., 1994). Conversely, no literature could be found suggesting that the prevalence rates are disparate because of the over-diagnosis of males. There is however, a body of literature suggesting that ADHD, in general, is over-diagnosed. One may reason that if ADHD is over-diagnosed, and males receive the label at a rate of 3 to 9 times that of females, then it may be a male

over-diagnosis, rather then a female under-diagnosis, causing the large prevalence difference.

## Comorbidity Considerations

Weinberg and Brumbach (1992) evaluated 100 consecutive referrals to the Behavioral Neurology Program of the Children's Medical Center of Dallas. Of these 100 children, 80% fulfilled the criteria for a diagnosis of ADHD and could have been labeled accordingly. However, "a closer examination of behavior (using a semi-structured, closed-end evaluation technique) and of higher cortical communicative functions, made it possible to provide a more exact diagnostic categorization" (p. 432). The authors found that all of the children were more accurately diagnosed with one of three disorders: affective illness, learning disability, or primary disorder of vigilance. The convenient label of ADHD would have kept physicians, schools and parents from understanding the multiplicity of clinical problems and led to improper treatment.

Weinberg and Brumbach (1992) suggested that when hyperactivity or misbehavior is apparent, the leading cause of ADHD symptoms is affective illness. When ADHD symptoms first appear with the onset of formal schooling, primary disorder of vigilance, or task-dependent attention disorder is the usual cause. The authors contended that these disorders are explainable as the result of "genetically based dysfunction of specific cerebral or brain stem areas" (p. 442). The authors concluded that ADHD is a "myth" with symptomology explainable by other specific causes.

Sabatino and Vance (1994) studied 55 male and 20 female children who had been previously diagnosed with ADHD. The children were referred to a

multidisciplinary clinic after medical or educational interventions had been ineffective. The clinic's comprehensive assessment included family and social histories, parent and teacher checklists of behavioral descriptors, and routine pediatric and pediatric-neurological examinations. Additional data included a routine battery of intelligence, visual and auditory perceptual, memory, academic achievement, and personality tests. Each member of the assessment team submitted a diagnosis, and the final diagnosis was agreed upon following a team meeting. Out of the 75 cases, the clinic team diagnosed 22 with ADHD undifferentiated, and 9 with ADHD with hyperactivity. Thus only 31 of the 75 children referred with a diagnosis of ADHD maintained that diagnoses. Of the 44 not confirmed as ADHD, 13 had information processing problems associated with learning disabilities, while 10 had either central auditory processing or receptive language difficulties. Thus, a third of the children had problems unassociated with ADHD. A major concern, was the number of children diagnosed with emotional problems. The primary symptoms were not hyperactivity, but rather impulsivity related to conduct disorder, oppositional defiant disorder, anxiety disorders, and separation, avoidance, and overanxious disorders. Although there was a high number of emotional disorders in the children, not a single child in the study, nor any of the families, was undergoing psychological or psychiatric therapy or family training. Sabatino and Vance concluded that ADHD is over-diagnosed, and questioned how many emotionally disturbed children are being treated inappropriately.

Cotungo (1993) studied 76 boys and 16 girls referred over a three-year period to a community mental health center for ADHD evaluations. The children had all been previously diagnosed with ADHD by pediatricians, physicians, or

mental health workers without the benefit of comprehensive evaluations. Typically, the diagnoses were rendered based on direct observations and behavior rating scales completed by parents and teachers. By contrast, the clinic's evaluation began by obtaining extensive medical, developmental, educational, familial, and social histories. Licensed psychologists then administered a comprehensive battery of 16 tests covering intellectual, cognitive, academic, personality, and behavioral domains. Primary and secondary diagnoses were determined at a case conference chaired by senior clinicians or supervisors. Despite the fact that ADHD was the primary reason for referral, only 22% of the sample received a primary diagnosis of ADHD, while an additional 37% received ADHD as a secondary diagnosis. Undifferentiated attention deficit disorder (UADD) was the primary diagnosis in 4% of the cases, and the secondary diagnosis in 20% of the sample. The authors found that the referral sources had consistently given an inordinate amount of weight to the symptoms of inattention, distractibility and overactivity to the exclusion of symptoms characteristic of other disorders. They attributed this to the exclusive use of observational and behavioral rating data and the exclusion of developmental histories.

When primary and secondary diagnoses were taken together, 59% of the sample received a diagnosis of ADHD and an additional 24% were diagnosed with UADD; thus 83% of the children demonstrated evidence of ADHD. There appeared to be a consensus about the presence of the symptoms but less clarity about the diagnostic process. The more comprehensive process relegated ADHD to a secondary status in 57% of the children receiving the diagnosis. The ADHD symptoms were accounted for by other primary sources

in 74% of the sample. Contugno (1994) suggested that for a substantial number of children who demonstrate inattention and hyperactivity, anxiety, mood, and other behavioral disturbances may better account for the symptoms than does ADHD. The author concluded, "this unwittingly increases the likelihood of misdiagnosis or at the very least overdiagnosis of ADHD and UADD" (p.342).

## Labeling Bias

Mulhern, Dworkin and Bernstein (1994) investigated the ability of parental concerns for their children's behavior to predict a diagnosis of ADHD. This retrospective study included 245 children referred for comprehensive pediatric evaluations for school problems. The types of concerns parents identified were categorized as inattentive, impulsive or hyperactive. These were compared to the children's final diagnosis to determine each symptom's specificity, sensitivity and predictive value. Significant school-related problems were diagnosed in 92% of the subjects, while only 38% received a diagnosis of ADHD. The authors concluded that parental concerns of ADHD symptoms identified many children without the disorder; however, the identified children did have other significant school-related problems.

Halperin, Matier, Bedi, Sharma, and Newcorn (1992) compared children with ADHD to non-ADHD psychiatric patients and normal controls on objective measures of inattention, hyperactivity, and impulsivity to determine the specificity of the symptoms to ADHD. Using 104 children from their outpatient clinic, they found the two patient groups to be indistinguishable from one another on measures of attention. The authors cautioned that the high proportion of non-ADHD children with inattention suggests that children with

other psychiatric diagnoses are "at risk to be mislabeled ADHD, simply because the ADHD diagnosis carries the name of the symptom domain" (p.194).

Desgranges, Desgranges, and Karsky (1995) reviewed patient records on all cases initiating treatment in their clinic in 1993. Of 375 cases reviewed, 119 came requesting an evaluation for ADHD or presenting with symptoms consistent with ADHD. Only 38% of these cases were confirmed as having ADHD. The authors suggested that without careful assessment, individuals with other problems could have been overlooked or misdiagnosed. Even more alarming, however, was the impact of preconceived diagnosis on treatment compliance. The group with confirmed ADHD and the group with other problems had similar treatment success rates, while the group seeking, but not getting, a diagnosis of ADHD had a large increase in client-initiated treatment terminations. The authors suggested that this group became frustrated because they were "so convinced that they knew the diagnosis, they did not process information and often did not even complete the evaluation when told other factors were being considered" (p.16). Desgranges et al. (p.16) concluded that ADHD has become an "overused catch-all used by schools, families, the public, and some medical / mental health professionals to explain/excuse a wide range of problems."

## Public Policy Influences

Since the implementation of Public Law 101-476, the effect of ADHD on academic performance has been recognized and requires classroom modifications be specified in an Individual Educational Plan (Wolraich & Baumgaertel, 1997). Cooper and Ideus (1995) suggested that receiving an ADHD label is no longer viewed as a stigma, but rather a source of

empowerment, providing a means by which individuals may secure special education and other additional funding to meet their educational, medical, and social needs. Sabatino and Vance (1994) evaluated school-related reasons for the referral of children who were inaccurately diagnosed with ADHD. The referrals from school personnel suggested that the children had ADHD; yet, the overlap between school-related reasons for referral and the 14 symptoms listed in the DSM-III-R for ADHD were extremely limited. Sabatino and Vance suggested that school personnel do not think in terms of 14 symptoms, but rather think about how a child has to be labeled in order to obtain services.

Barkley (1998) stated that physicians have recently become overly enthusiastic about encouraging adults to seek a diagnosis of ADHD. This rise in adult ADHD has roughly coincided with the advent of the Americans with Disabilities Act of 1990 (Public Law 101-336). According to this law, ADHD is considered a legitimate disability, requiring employers, educational institutions, and testing organizations to make reasonable accommodations for those who suffer the disorder. Given this environment, Barkely fears that overly liberal diagnostics will ultimately hurt the cause of protecting the rights of those legitimately afflicted.

## Oppositional Defiant Behavioral Confounds

Studies by Schachar, Sandburg, and Rutter (1986), Abikoff, Courtney, Pelham, and Koplewicz (1993), and Stevens, Quittner and Abikoff (1998), provide indirect evidence that ADHD may often be misdiagnosed, particularly among boys, as a result of the confounding effects of oppositional defiant disorder, and other disruptive forms of behavior.

Schachar et al. (1986) sampled children for their study by having teachers complete a Conners Teacher Rating Scale (CTRS; Conners, 1969) on 185 boys in their final year of infant school (North American grade 1 equivalent). The children were then grouped into three categories (high, medium, or low) according to their hyperactivity factor from the CTRS. Thirty-three children were then randomly selected from each category. Three different measures of the children's behaviors were taken over a several week period. First, a direct observation, time sampling procedure was used in which raters documented the classroom behaviors of the target children corresponding to the CTRS hyperactivity, inattentiveness, and behavior problem factors. Second, an 11-item Direct Observation Questionnaire (DOQ) was developed by the authors to assess behaviors similar to those from the direct observation and CTRS. These were completed by classroom teachers who were instructed to rate the children on their behaviors during the previous week. Third, observers and teachers completed a CTRS for each child.

Schachar et al. (1986) found considerable agreement between the three measures. Of particular interest was their finding of a "halo effect" whereby the presence of particular behaviors affected ratings of phenomenologically different behaviors (Guilford, 1954). They found that difficult relationships with peers and teachers increased teacher ratings of hyperactivity. There was, however, no such halo effect of hyperactive behaviors on teacher ratings of behavior problems. Children with poor relationships were more likely to be rated as hyperactive, regardless of observed activity level. Aggressive and defiant social interactions with teachers and peers also created a similar halo effect on rated inattentiveness. Children who exhibited aggressive or defiant

behaviors were more likely to be rated as inattentive compared to children who were not aggressive or defiant. Ironically, both positive interaction with peers and inattention were associated with inflated behavior problem ratings. Since teacher rating scales are relied upon, almost exclusively, to determine the prevalence rates of ADHD (Abikoff et al., 1993) and since hyperactivity and inattention, two of the three behavioral domains of ADHD, were shown to be overrated in the presence of aggressive and defiant behaviors, it is likely that the presence of such behaviors has contributed to the inflation of ADHD prevalence rates.

Abikoff et al. (1993) used an analogue design to study the impact of oppositional defiant disorder (ODD) and ADHD on teachers' ratings of children's behavior. One hundred and thirty-nine regular and special education elementary school teachers viewed ten-minute, videotaped segments of what they believed were children in a regular fourth grade classroom. The children were actually actors, with one male child in each tape engaging in behaviors reflective of either ADHD or ODD. After viewing each tape, the teachers rated the target child on a 73-item, 4-point scale comprised of the following: (a) items from the CTRS Hyperactivity factor and Hyperkinesis Index; (b) the IOWA Conners Aggression factor; (c) the IOWA Inattentive / Overactive factor; (d) verbatim descriptors of the items comprising the DSM-III-R ADHD and ODD diagnostic categories; and (e) four DSM-III-R conduct disorder symptoms. Like Schachar et al. (1986), Abikoff et al. found evidence of a unidirectional bias in teacher ratings. The teachers accurately rated ADHD behaviors when the child behaved accordingly; yet, when the child engaged in behaviors associated with ODD, they also rated that child as exhibiting ADHD behaviors. By contrast,

teachers accurately rated ODD behaviors, regardless of the child's activity level. The authors concluded that oppositional behaviors created a negative halo, resulting in elevated ADHD ratings.

Stevens, Quittner and Abikoff (1998) replicated Abikoff et al.'s (1993) design to examine the factors which might lead teachers to be biased in their ratings. First, they hypothesized that two factors in rating scales may affect teachers' abilities to distinguish ADHD from ODD: the degree of confounding symptoms across the two disorders and the degree to which the measures are behaviorally anchored. Second, they questioned whether the teachers' knowledge of ADHD would impact their ability to distinguish between the two behavior disorders. Third, they examined if teachers' professional involvement with ADHD would impact their ratings.

After viewing the same videotapes used by Abikoff et al. (1993) Stevens et al. (1998) had their 108 elementary school teachers complete two different rating scales: the CTRS-28, and the SNAP-IV. In addition, the teachers completed two questionnaires assessing their knowledge of and experience with ADHD. Similar to the findings of Abikoff et al., the authors found that the presence of oppositional behaviors exerted a unidirectional negative halo on teachers' ratings of hyperactivity and inattention; however, Abikoff et al. found evidence for a somewhat stronger halo effect. Stevens et al. also found that teachers were less biased when using the behaviorally anchored SNAP-IV than the more global CTRS-28. Teachers' professional experience with and knowledge of ADHD were found to have no impact on their ratings.

This unidirectional halo effect may also underlie findings from Prinz, Conner and Wilson (1981). The authors examined the relationship between

aggression and hyperactivity using 109 first-through third-grade children who were selected by their classroom teachers as being "disruptive and difficult to control" (p.194). These children, along with a group of control children, were rated by the teachers daily for twelve consecutive school days using the CTRS and the Daily Behavior Checklist (DBC). The DBC, a measure specifically designed for the study, is a checklist describing specific manifestations of hyperactivity or aggression directed at either people or objects. The results of the study showed that, for children rated as hyperactive, aggressive behavior was the highest among those who exhibited the highest levels of hyperactivity. In addition, teachers more often recorded hyperactive behaviors on days when they reported aggressive behaviors, whereas the conditional probability of aggressive behaviors was not associated with the reported occurrence of hyperactive behaviors.

This unidirectional bias could also help explain the findings of Pelham, Evans, Gnagy and Greenslide (1992) and Pelham, Gnagy, et al. (1992). While studying the internal consistency and factor structure of their Disruptive Behavior Disorders Scale (DBD) they found that ODD symptoms generally predicted the presence of ADHD as well as did the presence of ADHD symptoms, but not visa versa.

The unidirectional negative halo of oppositional defiant behaviors on ratings of ADHD may offer a partial explanation for the higher prevalence of males with ADHD. Gaub and Carlson's meta-analysis (1997) showed that males with ADHD show higher rates of externalizing behaviors than girls with ADHD. This pattern results in more boys displaying disruptive behaviors in structured settings, such as school, possibly leading to higher rates of referral.

Once referred, teachers complete rating scales on these disruptive boys and, as we have seen, teachers are subject to biased ratings due to the negative halo effect. Research, which arguably has also been influenced by biasing effects, has routinely demonstrated that, regardless of the presence of ADHD, girls of all ages are less hyperactive and have fewer attention problems than same-aged boys (Achenbach, 1991; Bauermeister, 1992; Brito, Pinto, & Lins, 1995; Conners, 1994; Goyette, Conners & Ulrich, 1978; Trites, Blouin, & Laprade, 1980). Females also show fewer externalizing problems (Achenbach, 1991; Bauermeister, 1992; Conners, 1994; Eme, 1992; Garb & Carlson, 1997; Lahey, 1994; Zoccolillo, 1993) and experience more social acceptance and less peer rejection than boys (Arnold, 1996). In essence, the average boy is more likely to present a management problem and be referred for services than the average girl.

McGee and Feehan (1991) presented evidence that teacher ratings show more strongly pronounced sex differences for ADHD behaviors than do parent ratings. For example, McGee, Williams and Silva (1987) found a boy-girl ratio of 7:1 for teacher-identified ADHD, which dropped to 1.5:1 with parent ratings. The Ontario Child Health Study (Szatmari, Offord, & Boyle, 1989) found that the sex difference in prevalence rates of ADHD for children ages 4 to 11 was almost entirely accounted for by teacher ratings. Studying American and Puerto Rican children and adolescents, Achenbach (1990) had parents complete the Child Behavior Checklist (CBCL) and teachers the Teacher Report Form (TRF). For parent reports, analyses of total scores revealed no differences between the genders from either culture. Teachers, by contrast, rated males from both cultures as higher on ADHD measures than females. Similarly, Breen

and Altepeter (1990) evaluated parent and teacher reports of behavior management problems on children with ADHD. Once again, sex differences on these ratings were restricted to teacher reports. In the article describing their meta-analysis, Gaub and Carlson (1997) essentially agreed with McGee and Feehan's contention that the gender-based prevalence difference in ADHD is largely due to teacher ratings: "While both parents and teachers rated boys with ADHD as being more deviant than girls with ADHD on ratings of inattention and hyperactivity, the average effect size differences between genders were significantly larger for teacher than for parent evaluations" (p.1043).

Other than the fact that parents and teachers view children in different settings and that they may have different "anchors" for their ratings (Gaub & Carlson, 1997), little has been suggested to account for their disparate gender ratings on ADHD scales. It is quite possible that just like the psychiatrists who read identical descriptors, then diagnose more females than males as exhibiting histrionic personality disorder (Loring & Powell, 1988), teachers may also be highly influenced by their expectations which tell them that boys are more disruptive management problems than girls. Stevens (1980) showed how perceived socioeconomic status and race can influence teacher ratings. Shown three-minute silent films in which only race and socioeconomic status varied, teachers attributed significantly fewer negative behaviors to middle-class children than to perceived lower socioeconomic children, and rated African-American children as more deviant than Mexican-American children. The attribution of behavioral traits has repeatedly been shown to be influenced by individual characteristics such as attractiveness, gender, socioeconomic status, and race (Stevens, 1980).

In summary, research has consistently shown that males are diagnosed with ADHD more frequently than females. The evidence suggests that this may be the result of bias, and not a reflection of the actual base rate of the disorder. Research also suggests that ADHD, in general, may be over-diagnosed, and that biasing effects may contribute to the over-diagnosis. Abikoff, Courtney, Pelham and Koplewicz (1993), and Stevens, Quittner and Abikoff (1998) found that oppositional defiant behaviors created a "negative halo" on teacher ratings of ADHD, which spuriously inflated the perceived occurrence of ADHD behaviors. This "negative halo effect" was found to be unidirectional, as the presence of ADHD behaviors had no impact on teacher ratings of oppositional defiant behaviors.

## Present Study

The generalizability of the two studies (Abikoff, Courtney, Pelham, & Koplewicz, 1993; Stevens, Quittner & Abikoff,1998) which have directly explored the "unidirectional negative halo effect" are limited by their reliance on a single, male target-child in their analogue tapes. The current study replicated the two earlier studies, with a new set of tapes using the same scripts, and included tapes with female targets. Teacher ratings were expected to be influenced by three factors: (a) the type of behaviors portrayed by the target child on the analogue tapes (ADHD, ODD or normal), (b) characteristics of the rating instruments (global vs. descriptive ratings), and (c) gender of the target child. It was hypothesized that both male and female targets displaying ODD behaviors would be accurately accurately rated as exhibiting ODD behaviors. However, it was expected that teachers would inaccurately rate males displaying ODD behaviors as also displaying ADHD behaviors. Conversely, it

was expected that fewer inaccuracies would be found in teacher ratings of females displaying ODD behaviors. In addition, the degree of inaccurate reporting was expected to vary as a function of the type of ADHD rating instrument employed. Measurements utilizing more specific operational definitions and requiring more descriptive (e.g., "fidgets with hands or squirms with feet") as opposed to global (e.g., "childish and immature") judgments, were expected yield fewer inaccuracies.

## Methodological Considerations

The two previous studies that directly evaluated the unidirectional negative halo effect of oppositional defiant behaviors on teachers' ratings of ADHD (Abikoff, Courtney, Pelham, & Koplewicz, 1993; Stevens, Quittner & Abikoff,1998) provided evidence for the presence of the unidirectional negative halo effect by showing that a child portraying ODD-symptomatic behaviors and a child portraying ADHD-symptomatic behaviors did not significantly differ on teacher-completed ADHD rating scales. Finding such results critically depended upon the total number and frequency of ADHD behaviors exhibited by the children on both tapes.

Guilford (1954) suggested that a halo effect existed when the presence of particular behaviors affect ratings of phenomenologically different behaviors. Methodologically, testing for the presence of a halo effect may be done in one of two ways. The first method requires three conditions: (a) the presence of two individuals, one of whom is exhibiting halo-suspected behaviors, such as oppositionality and defiance, while the second individual is either not exhibiting, or is exhibiting a very low level of the halo-suspected behaviors; (b) the presence of a phenomenologically different behavior, such as hyperactivity, that

the two individuals are exhibiting at or near the same level; and (c) finding that the two individuals are rated as exhibiting significantly different amounts of the phenomenologically different behavior. Since the two individuals are exhibiting different amounts of the halo-suspected behavior, and the same amount of the phenomenologically different behavior, the differences found on rating scales would logically be attributed to the effect of the halo-suspected behavior.

The second method of testing for the presence of a halo effect also requires three conditions: (a) the presence of two individuals, one of whom is exhibiting halo-suspected behaviors, such as oppositionality and defiance, while the second is either not exhibiting, or exhibiting a very low level, of the halo-suspected behaviors; (b) the presence of a phenomenologically different behavior, such as hyperactivity, that one individual is exhibiting, while the second individual is either not exhibiting the behavior, or is exhibiting the behavior at a very low level; and (c) showing that the two individuals are rated as exhibiting the same level of the phenomenologically different behavior. Since the two individuals were exhibiting different levels of the phenomenologically different behavior, the lack of difference found between the two would logically be attributed to the effect of the halo-suspecting behavior.

Abikoff et al. (1993) and Stevens et al. (1998) utilized this second approach in their studies; likewise, the current study utilized this method to test for the presence of the unidirectional negative halo effect. However, a second viable approach, which utilized the first method outlined above, was to compare the pathology tapes (ADHD/ODD) to the normal tapes. Since the tape validation process in the current study found that the ODD tapes and the normal tapes exhibited similar amounts of ADHD-type behaviors, and that the ADHD

tapes and the normal tapes exhibited similar amounts of ODD-type behaviors, it was appropriate for the present study to utilyze the normal tapes to test for the presence of the halo effect. Specifically, the ODD and normal tapes were compared on the four ADHD rating scales, and the ADHD and normal tapes were compared on the two ODD measures.

It was hypothesized that teachers would inaccurately rate the male ODD tape as exhibiting significantly more ADHD-type behaviors, as compared to the female ODD tape, and as compared to the male and female normal tapes. Given the previous findings of a unidirectional halo effect, it was also hypothesized that no such bias would be found when comparing the ADHD and normal tapes on ODD rating scales.

# CHAPTER II

# METHOD

## Participants

The initial sample consisted of 47 special education and 44 regular education, kindergarten - sixth grade school teachers, drawn from small cities and rural communities in the upper Midwest and one large city from the Pacific Northwest. Participants were recruited with the assistance of school principals and a special education coordinator. Teachers were also obtained through a Masters' degree training program at a state university. A total of 11 teachers was dropped from the final analysis: three failed to return their questionnaires, six failed to adequately complete their questionnaires, and two acknowledged having known a child on the tapes. The final sample consisted of 40 special education and 40 regular education teachers. A total of 21 schools was represented in the study. Based on the population parameters suggeted by Offord et al. (1987), which defined urban areas as those with a population of more than 25,000 people, the sample contained 59 urban and 21 rural teachers. The majority of the teachers was women (95%), and the sample had considerable teaching experience ($\underline{M}$ = 14.37 years, $\underline{SD}$ = 9.59). The participants viewed the tapes at their respective schools in groups ranging from 1 to 10 subjects per screening. The rural teachers received their questionnaire packets through inter-school mail, then participated in the screenings via live-interactive television. All packets remained sealed until participants were instructed to open them prior to viewing the first tape. Following the screenings,

materials were either directly collected or returned to the author through inter-school mail.

## Procedure

After signing an informed consent sheet (Appendix A), the teachers viewed two, 9-minute videotaped segments of fourth grade children involved in a classroom exercise. Prior to viewing the tapes they were given the following instructions:

> You will be presented with two videotapes showing children in a regular fourth grade classroom containing children with mixed learning levels. Special attention is given to one child in each tape who will be pointed out as the tape begins. You can think of each tape as a ten minute 'slice of life in the school day of a child' where the children are expected to be doing their individual seat work. Your task is to watch the targeted youngster and, when the tape is over, to complete the questionnaires contained in the packet handed to you. As you watch the tapes it is important that you know that our prime interest is in your judgments of the child's behavior, rather than your perceptions of the teacher depicted in the tape. These are composite tapes, made to capture the child, and no effort was made to accurately reflect the teacher's skills (Abikoff, et al., 1993, p. 521).

Teachers were also instructed not to look at their questionnaires until the end of each tape. Two questionnaires were handed out, one for each tape viewed. When teachers asked if they should extrapolate from the behaviors observed in order to answer questions which addressed behaviors not readily observable from the tapes, they were given the following instructions:

> This is a typical problem you teachers face when you are asked to complete these kinds of questionnaire about an actual child in your classroom. The forms never quite fit the child being rated. All we can say is that you should solve the problem today in the same way that you would do it in the real situation" (Abikoff et al., 1993, p.522-523).

Though led to believe they were viewing tapes of actual classrooms, they were viewing a setting in which the children and teacher were actors following prepared scripts. Three tapes depicted a boy exhibiting behaviors characteristic of either (a) pure ADHD, (b) pure ODD, or (c) normal behaviors. Likewise, three tapes depicted a girl following the same script and showing the same behaviors. To the extent possible, the tapes were created to ensure that children on the pure ADHD tapes did not engage in ODD behaviors. In addition, the rate of ADHD behaviors displayed by children on the pure ODD tapes were not deviant, and were equivalent to the rate of ADHD behaviors displayed on the normal tapes.

A between-subjects design was employed. Teachers were assigned to one of four viewing conditions in which they watched two tapes in succession. One tape was the control tape in which the target child was not exhibiting pathological behaviors, while the second tape portrayed either ADHD or ODD behaviors. The gender of the target child varied. The four viewing conditions were as follows: (a) male ADHD/male normal control, (b) male ODD/male normal control, (c) female ADHD/female normal control, and (d) female ODD/female normal control. This study contained two independent variables: type of tape (ODD vs. ADHD) and gender of the target child.

Two extraneous variables that could have impacted the study, order of tape presentation (pathology tape first vs. normal tape first), and type of teacher (regular vs. special education), were addressed in the assignment of teachers to the four viewing conditions. Within the four conditions, order of tape presentation and type of teacher were successfully counterbalanced. Ten special education and 10 regular education teachers were assigned to each separate viewing condition. Half of the teachers in each group viewed the pathology tape first, and half viewed the normal tape first. In addition, type of teacher was included in the counterbalancing of the order of tape presentation, such that within each of the four conditions, five special education and five regular education teachers viewed the pathology tape first, while five special education and five regular education teachers viewed the normal tape first.

## Dependent Variables

The purposes of the current study were two-fold: replicate previous findings (Abikoff, et al., 1993; Stevens et al., 1998) using different tapes, and to determine the extent to which the negative halo effect impacts females. Therefore, similar dependent measures used in the previous studies were maintained, except where theoretical grounds suggested the use of different measures.

### Conners' Teacher Rating Scales - Revised: Short

Both previous studies used the Conners' Teacher Rating Scale-28 (CTRS-28; Conners, 1973), a shortened version of the original, 39-item Conners' Teacher Rating Scale (CTRS; Conners, 1969). The Conners' Rating Scales (CRS) were developed to aid in the identification of hyperkinetic children and to evaluate the treatment effectiveness of medication (Goyotte,

Conners, & Ulrich, 1978). Due to their low cost and brevity, the Conners' Scales have become some of the most widely used scales (Barkley, 1998), with over 1.5 million rating scales used per year (Conners, 1997). The scales have been shown to discriminate between normal and hyperactive children, and to be sensitive to drug treatment effects (Goyotte et al., 1978). The CRS were formally published for the first time in 1989, and revised versions followed in 1997 (Conners' Rating Scales-Revised; CRS-R). The goals of the revisions were threefold: to recognize and incorporate aspects of the DSM-IV, include new normative data, and introduce adolescent self-report scales (Conners, 1997).

The Conners' Teacher Rating Scale-Revised: Short Form (CTRS-R:S) dropped four of the items from the original CTRS-28. Loney and Milich had used the CTRS-28 to develop the IOWA Conners' Teacher Rating Scale (Loney & Milich, 1982). Although Abikoff et al. (1993) used the IOWA Conners' in their study due to its "purer" measures of inattention, aggression, and activity level compared to the CTRS-28 factors, the present study used the CTRS-R:S due to its' improved psychometric properties relative to the CTRS-28. As a result, the IOWA Conners, which consists of ten items from the CTRS-28, was not utilized in the current study. It should also be noted that Stevens et al. (1998) used the CTRS-28 in their study, which would have allowed for the calculation of the IOWA Conners', yet they chose not to use the measure.

The Conners' Teacher Rating Scale-Revised: Short Form (CTRS-R:S) is a 28 item questionnaire in which each item is answered on a 4-point scale (Not true at all = 0; Just a little true = 1; Pretty much true = 2; and Very much true = 3). The measure has four subscales: Oppositional, Hyperactivity, Cognitive

Problems, and an ADHD Index. Abikoff et al. (1993), using the CTRS-28, only utilized the Hyperactivity Factor and Hyperkinesis Index, while Stevens et al. (1998) utilized three of the four scales from the CTRS-28, omitting only the Cognitive Problems Factor. With no theoretical justification to drop or add scales, the current study employed the Oppositional and Hyperactivity Factors and the ADHD Index from the CTRS-R:S.

The CTRS-R:S has good psychometric properties. The Oppositional, Hyperactivity and ADHD Index have 6-8 week test-retest reliabilities of .84, .72, and .84, respectively. The internal reliability coefficients fluctuate depending upon the child's gender and age. The coefficients range from a high of .99 on the Hyperactivity factor for 15-17 year-old females, to a low of .78 on the Oppositional Factor for 15-17 year-old females (Conners, 1997).

Disruptive Behaviors Disorders Rating Scale

Seeking to employ a measure utilizing specific operational definitions, and requiring more descriptive, as opposed to global, judgments contained in the Conners' Rating Scales, Stevens et al. (1998) included the SNAP-IV Rating Scale. Swanson, Nolan and Pelham developed the original SNAP Rating Scale (Atkins, Pelham, & Licht, 1985) to remedy the problem that other rating scales at the time did not include as items the statements listed as behavioral descriptors of the disorders found in the DSM (Pelham, Gnagy, Greenslade & Milich, 1992). The original SNAP listed the DSM-III symptoms of ADD in a 4-point rating scale format. Norms were gathered, and the scale was employed in numerous studies (Pelham et al. 1992). Those norms, however, may no longer be valid, given the numerous changes which took place with the transition to the DSM-III-R and DSM-IV. Pelham et al. attempted to remedy this problem with

the creation of the Disruptive Behaviors Disorder Rating Scale (DBD). Using the same 4-point response format of the CRS, the DBD Rating Scale includes, as nearly as possible, the exact wording of the 36 descriptors of all three of the disruptive behavior disorder categories from the DSM-III-R: attention deficit, conduct, and oppositional defiant disorder. The scale was normed and validated in two large national studies (Pelham, Evans, Gnagy, & Greenslade, 1992; and Pelham, Gnagy, Greenslade, & Milich, 1992). Factor analysis revealed three factors. The first factor, termed Oppositional/Defiant, consisted of eight DSM-III-R ODD items, two DSM-III-R CD items, and two DSM-III-R ADHD items. The second factor, Inattention, was comprised of nine ADHD items. The final factor, Impulsivity/Overactivity, consisted of nine ADHD items and one ODD item. Pelham, Gnagy, Greenslade and Milich (1992) concluded that the DBD Rating Scale was useful to aid in the diagnosis of ADHD and ODD, but not CD.

Stevens et al. (1998) used the SNAP-IV Rating Scale, which to date had not been validated. In addition, use of the scale was logically problematic because the child actors being rated depicted ADHD and ODD behaviors based on the DSM-III-R conceptualization. Thus, in keeping with Stevens et al.'s intentions of using operationally defined, descriptive judgments, while maintaining congruence with the behavioral symptoms portrayed, and desiring a validated measure, the current study employed the 24 DBD Rating Scale questions measuring the constructs of inattention, impulsivity/overactivity, and oppositional defiance, rather than the SNAP-IV which measured similar constructs.

## Demographic Information

The questionnaire also collected the following demographic information: (a) type of teacher (i.e., regular or special education); (b) years of teaching experience; (c) gender; (d) age; and (e) location of school (i.e., rural or urban). Urban areas were defined as those with a population of more than 25,000. Rural areas included both small urban areas (population of 3,000 - 25,000) and rural areas (population less than 3,000) (Offord et al., 1987). Stevens et al. (1998) found that neither "knowledge of ADHD" nor "experience with ADHD" affected the accuracy of teacher ratings; consequently, these variables were not addressed in the current study.

## Summary of Dependent Variables

In summary, the teachers in the present study were given a 55-item questionnaire consisting of 23 items from the CTRS-R:S, 24 items from the DBD Rating Scale, and 8 demographic questions. From this questionnaire, six dependent measures were calculated, four measuring ADHD behaviors and two measuring ODD behaviors. The two ODD measures were as follows: (a) the Oppositional Factor from the CTRS-R:S, and (b) the Oppositional/Defiant Factor from the DBD. The four measures of ADHD included: (a) the Hyperactivity Factor from the CTRS-R:S, (b) the ADHD Index from the CTRS-R:S, (c) the Inattention Factor from the DBD, and (d) the Impulsivity/Overactivity Factor from the DBD.

## Independent Variables

There were two independent measures in this study: (a) videotapes portraying "pure" ADHD or "pure" ODD behaviors, and (b) gender of the target child on the videotapes . The videotapes were developed to meet two goals.

First, the target child was to exhibit behaviors, in both type and frequency, that portrayed ADHD, ODD, and normal children. On the ADHD and ODD tapes, the target child's behaviors were to clearly distinguish him or her from the other children on diagnostically relevant criteria. Second, to assure that raters were responding to the target child's behavior and not others' responses to the target child, the behavior of the teacher and other children in the classroom were controlled.

Child actors were recruited from a private school, and parents were required to sign an informed consent form (Appendix B) before the children could participate. Both the children and the teacher-actor were paid for their participation. The scenes were shot in an actual fourth grade classroom, where the children were involved in a lesson which required both following instructions and working independently. The scripts that the actors followed were transcribed directly from the tapes validated by Abikoff et al. (1993) (Appendix C, Appendix D, Appendix E). The researchers wrote these scripts to portray the relative behavior rates of ADHD, ODD and normal children. These behaviors and rates were based on classroom observation data collected using two observation codes which had been shown to differentiate pathological and normal children: the Revised Stony Brook Observation Code (Abikoff, Gittelman-Klein, & Klein, 1977), and the Classroom Observations of Conduct and Attention Deficit Disorder (COCADD; Atkins, Pelham, & Licht, 1985, 1988, 1989).

Revised Stony Brook Observation Code

The tapes in the present study were validated using 10 paid graduate students, who were blind to the type of tape they were rating. The students

were trained on the Revised Stony Brook Observation Code scoring system (Abikoff & Gittelman, 1985). This observation procedure employs a 15-second time sampling method in which the target child is rated on 11 categories of behavior: (a) interference, (b) off-task, (c) minor motor movement, (d) gross motor movements-standing, (e) gross motor movements-vigorous, (f) out of chair, (g) physical aggression, (h) verbal aggression towards another child, (i) verbal aggression towards the teacher, (j) noncompliance, and (k) solicitation of teacher. The first six categories were combined to form a composite hyperactivity score, and the next four categories were combined to form a composite oppositional score. Solicitation was not used due to the category's non-specific nature; it is neither unique to ADHD nor ODD.

The results from the validation process indicate that the manipulations were successful. As shown in table 1, the mean Hyperactivity Composite scores for the two ADHD tapes were approximately 4 times greater than the control tapes, and 3 1/2 times greater than the ODD tapes. Again, the mean Oppositional Composite scores for the two ODD tapes were approximately 7 times greater than the ADHD tapes, and over 15 times greater than the normal tapes. In addition, the behavior rates within each tape category (ADHD, ODD, normal) were similar between males and females.

For comparison purposes, Table 1 also presents the Revised Stony Brook validation results from Abikoff et al. (1993). The Hyperactivity Composite scores from the ADHD tapes were similar between the tapes from the Abikoff et al. study ($\underline{M}$ = 54), and the male ($\underline{M}$ = 49.0) and female ($\underline{M}$ = 52.2) tapes from the present study. However, the Hyperactivity Composite scores were considerably higher in the Abikoff et al. ODD tapes ($\underline{M}$ = 25), compared to the

Table 1. Mean Behavior Rates for Each Tape[a]

| Behavior | Type of Tape | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Normal | | | ADHD | | | ODD | | |
| | Male | Female | Abikoff | Male | Female | Abikoff | Male | Female | Abikoff |
| Hyperactivity[b] | 13.00 | 12.80 | 24.00 | 49.00 | 52.20 | 54.00 | 14.80 | 14.80 | 25.00 |
| Off-Task | 1.60 | 3.00 | * | 12.40 | 8.60 | * | 2.20 | 2.80 | * |
| Interference | 4.80 | 4.80 | * | 18.00 | 17.20 | * | 6.80 | 6.60 | * |
| M.M. Mvts. | 6.60 | 5.00 | * | 9.60 | 13.80 | * | 4.80 | 3.80 | * |
| G.M. Vig. | 0.00 | 0.00 | * | 1.00 | 0.80 | * | 0.00 | 0.00 | * |
| G.M. Stand | 0.00 | 0.00 | * | 3.00 | 5.60 | * | 1.00 | 1.00 | * |
| Out of Chair | 0.00 | 0.00 | * | 5.00 | 6.00 | * | 0.00 | 0.60 | * |
| Oppositional[c] | 0.20 | 0.80 | 0.00 | 0.80 | 1.20 | 1.00 | 8.00 | 7.60 | 12.00 |
| Non-comp. | 0.20 | 1.00 | 0.00 | 0.40 | 0.20 | 1.00 | 0.40 | 0.60 | 4.00 |
| Phys. agres. | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 1.20 | 2.00 |
| V. agrs. chld. | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 2.80 | 2.00 | 2.00 |
| V. agrs. tchr. | 0.00 | 0.00 | 0.00 | 0.40 | 1.00 | 0.00 | 3.80 | 3.80 | 4.00 |

Note. The data in columns 3, 6 and 9 are from Teachers' Ratings of Disruptive Behaviors: The Influence of Halo Effects, by H. Abikoff, M. Courtney, W. E. Pelham, & H. Koplewicz, 1993, Journal of Abnormal Child Psychology, 21, p. 523. Copyright 1993 by Plenum Publishing Corporation. Adapted with permission.

Note. * = data not available

[a]Based on Revised Stony Brook Observation Code (Abikoff & Gittelman, 1985).

ADHD = attention-deficit hyperactivity disorder; ODD = oppositional defiant disorder.

[b]The sum of the code categories: off-task, interference, minor-motor movements (M. M. mvts.), gross-motor: vigorous (G. M. Vig.), gross-motor: standing (G. M. Standing), and out-of-chair.

[c]The sum of the code categories: non-compliance, physical aggression, verbal aggression towards a child (V. agrs. child), verbal aggression towards teacher (V. agrs. tchr.).

male (M = 14.8) and female (M = 14.8) ODD tapes in the present study. This same pattern also held true on the normal tapes, where the Hyperactivity Composite scores were nearly twice as high in the Abikoff et al. study (M = 24), compared to the male (M = 13.0) and female (M = 12.8) tapes in the present study.

On the ADHD tapes, the Oppositional Composite scores from Abikoff et al. were similar to the male and female scores in the present study (M = 1.0, 0.8, and 1.2 respectively). Likewise, on the normal tapes, the Oppositional Composite scores from Abikoff et al. and the present study were similar (M = 0.0, 0.2, and 0.8 respectively). However, on the ODD tapes, the Oppositional Composite scores were higher in the Abikoff et al. study (M = 12.0) compared to the male (M = 8.0) and female (M = 7.6) tapes in the present study.

Interobserver Reliability

Interobserver reliability measures were computed for each of the ten categories from the Revised Stony Brook Observation Code. The author always served as the "standard" in determining interobserver reliability. For each observation category, phi coefficients were determined as a measure of interval reliability (Gelfand & Hartman, 1975). The coefficients ranged from .31 to 1.0, with a mean $\Phi$ = .83 for all 10 categories. Table 2 presents the coefficients obtained with each observer for each category. These coefficients are similar to those obtained in the original validation study for the Revised Stony Brook Observation Code (Abikoff, Gittelman-Klein & Klein,1977), where $\Phi$ ranged from .34 - .93, with a mean $\Phi$ = .76 for all categories.

Table 2. Phi Coefficients Between the Interval Scores of the Standard and Other Observers on the Revised Stony Brook Observation Code[a]

| Category | Observer (N = 228 intervals) | | | | | Mean |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | |
| Off-Task | .66 | .82 | .75 | .68 | .76 | .73 |
| Interference | .74 | .82 | .80 | .83 | .65 | .77 |
| M.M. Mvts. | .72 | .74 | .72 | .68 | .65 | .70 |
| G.M. Vig. | 1.00 | 1.00 | 1.00 | 1.00 | .71 | .94 |
| G.M. Stand. | .95 | .85 | .90 | .95 | .90 | .91 |
| Out of Chair | .96 | 1.00 | .96 | .86 | .89 | .93 |
| Non-comp. | .50 | .71 | .40 | .81 | .31 | .55 |
| Phys. agres. | 1.00 | 1.00 | .81 | 1.00 | 1.00 | .96 |
| V. agrs. chld. | 1.00 | 1.00 | .80 | 1.00 | .89 | .94 |
| V. agrs. tchr. | .90 | .95 | .95 | .89 | .84 | .91 |

Note. [a]Abikoff & Gittelman, 1985.

M. M. Mvts. = minor-motor movements, G.M. Vig. = gross-motor: vigorous, G.M. Stand. = gross-motor: standing, Non-comp. = non-compliance, Phys.agres. = physical aggression, V. agrs. chld. = verbal aggression towards a child, V. agrs. tchr. = verbal aggression towards teacher.

All entries, $p < .001$.

## COCADD

Due to the importance of the independent measure, and in keeping with the intention of replicating Abikoff et al. (1993) the study utilized a second check on the external validity of the tapes: a modified version of the COCADD scoring

system (Carlson, Pelham, Milich, & Dixon, 1992) (Appendix F). Use of the COCADD proved somewhat problematic, as the system was designed to render an overall disruptive behavior summary score, not to differentiate between ADHD and ODD behaviors. The observation procedure entails recording the first disruptive behavior that occurs during each 15-second interval, regardless of the type of behavior; consequently, it does not yield an accurate measure of the frequency of different types of disruptive behaviors. In addition, the Destruction of Property category combines elements of both ADHD and ODD behaviors.

To remedy these problems, three modifications to the scoring procedure were recommended by one of the code's authors, E. M. Gnagy (personal communication, February 18, 2000) First, based on operational definitions, the eight categories were logically divided into two classifications, one reflecting ADHD-type behaviors and the other reflecting ODD-type behaviors. Second, the recording procedure was modified so that the first behavior from each classification was noted during each 15-second interval. Third, the Destruction of Property category was divided into two groups: the first, reflecting the actual destruction of property, was placed into the ODD classification; while the second, reflecting the inappropriate use of property, was placed into the ADHD classification (Appendix G).

The graduate students were trained on this revised version of the modified COCADD system. This system used a fifteen-second time sampling procedure rating the occurrence of eight disruptive behavior categories: (a) physical aggression/intrusion, (b) verbal abuse/teasing, (c) destruction of property, (d) cheating, (e) verbal intrusion, (f) talking to self, (g) leaving seat, and

(h) inappropriate use of property. Following each fifteen-second interval, there was an additional five-second period during which the the child was rated as being either on or off-task. The first four categories (physical aggression/intrusion, verbal abuse/teasing, destruction property, and cheating) were combined as an aggregate measure of ODD behaviors, while the second four categories (verbal intrusion, talking to self, leaving seat, and inappropriate use of property) were combined to derive a composite measure of ADHD behaviors. Although each 15-second interval was followed by a five-second period during which the observers rated the child as being on or off-task, the COCADD does not differentiate what off-task behaviors would be considered ADHD vs. ODD in nature; consequently, the off-task measure was not used.

Once again, the results from the validation process indicate that our manipulations were successful. As shown in Table 3, the percentage of 15-second intervals during which ADHD behaviors occurred were approximately 3 times greater on the ADHD tapes than they were on the ODD tapes, and 4 1/2 times greater on the ADHD tapes than they were on the normal tapes. The percentage of intervals during which ODD behaviors occurred were approximately 6 times greater on the ODD tapes compared to the ADHD tapes, and ODD behaviors were nonexistent on the normal tapes. In addition, the behavior rates within each tape category (ADHD, ODD, normal) were similar between the genders.

For comparison purposes, Table 2 also presents the modified COCADD validation results from Abikoff et al. (1993). The ADHD composite scores from both the male and female ADHD tapes in the present study indicated that ADHD-type behaviors occurred in considerably more of the 15-second intervals

compared to the ADHD tapes from Abikoff et al. (74.0% & 65.0% vs. 50. 0%). Likewise, the percentage of intervals containing ADHD behaviors in both the male (14.3%) and female (15.7%) normal tapes in the present study were higher than the normal tapes from Abikoff et al. (10%.). This pattern was reversed on the ODD tapes, where the ADHD composite rates for both the males and females in the present study indicated that ADHD-type behaviors occurred in fewer intervals compared to the ODD tapes from Abikoff et al. (19.0% & 17.0% vs. 29.0%).

The percentage of intervals during which ODD-type behaviors occurred on the ADHD tapes were similar for both the male (2.9%) and female (4.3%) tapes in the present study, and the ADHD tape from Abikoff et al. (1993) (3.3%). However, for the ODD tapes, the ODD composite rates from Abikoff et al. were considerably higher (35%) than both the male (21.7%) and female (21.4%) rates in the present study. On the normal tapes, no ODD behaviors were recorded in the present study or in Abikoff et al.

Table 3. Percentage of 20-second Intervals During Which Disruptive Behaviors Occurred[a]

| Behavior | Type of Tape | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Normal | | | ADHD | | | ODD | | |
| | Male | Female | Abikoff | Male | Female | Abikoff | Male | Female | Abikoff |
| ADHD-Composite[b] | 14.3 | 15.7 | 10.0 | 74.0 | 65.0 | 50.0 | 19.0 | 17.0 | 29.1 |
| ODD-Composite[c] | 0.0 | 0.0 | 0.0 | 2.9 | 4.3 | 3.3 | 21.7 | 21.4 | 35.0 |

Note. The data in column 3, 6 and 9 are from Teachers' Ratings of Disruptive Behaviors: The Influence of Halo Effects, by H. Abikoff, M. Courtney, W. E. Pelham, & H. Koplewicz, 1993, Journal of Abnormal Child Psychology, 21, p. 523. Copyright 1993 by Plenum Publishing Corporation. Adapted with permission.

[a]Based on modified version of the Classroom Observations of Conduct and Attention-Deficit Disorder code (Carlson, Pelham, Milich, & Dixon, 1992). ADHD = attention deficit hyperactivity disorder; ODD = oppositional defiant disorder.

[b]The sum of code categories: verbal intrusion, talking to self, leaving seat, and inappropriate use of materials, .

[c]The sum of code categories: physical aggression/intrusion, verbal abuse/teasing, destruction of property, and cheating

# CHAPTER III

# RESULTS

## Preliminary Analyses

There were no systematic differences in teachers' characteristics across the four conditions (ADHD/male; ADHD/female; ODD/male; ODD/female). A Chi-square test indicated that the number of urban and rural teachers were similar across the conditions, $\chi^2$ (3, $N$ = 80) = 1.23, $p > .05$. A one-way analysis of variance (ANOVA) revealed no significant age differences between the four conditions, $F$ (3, 76) = 1.52, $p > .05$. In addition, a one-way ANOVA indicated that the four groups did not significantly differ on years of teaching experience, $F$ (3, 76) = 1.87, $p > .05$.

Using each of the six dependent measures, Pearson product-moment correlations were computed between teachers' ratings on the pathology tapes and their ratings on the control tapes. No significant correlations were found. The obtained correlations ranged from a high of $r = .13$, $p = .25$, to a low of $r = .04$, $p = .71$. Due to a lack of significant correlations between teachers' scores on the control and pathology tapes, ratings of the control tapes were not used as covariates, as they had been in previous studies (Abikoff, Courtney, Pelham, & Koplewicz, 1993; Stevens, Quittner, & Abikoff, 1998).

As anticipated, paired $t$ tests indicated that teachers were able to differentiate between the normal and the pathology tapes on rating scales purporting to measure the given pathology (see Table 4). On the Conners'

Table 4. Teachers' Mean CTRS-R:S and DBD Ratings for the Normal, ADHD, and ODD Tapes

| Scale | Type of Tape | | | |
|---|---|---|---|---|
| ADHD Measures | Normal (N = 40) | | ADHD (N=40) | |
| | M | SD | M | SD |
| CTRS-R:S | | | | |
| HYP | 0.62 | 0.93 | 12.67[a] | 3.94 |
| ADHD | 2.22 | 2.95 | 25.82[b] | 6.22 |
| DBD | | | | |
| IO | 0.67 | 1.46 | 15.82[c] | 5.48 |
| IN | 1.12 | 2.29 | 16.15[d] | 5.15 |
| ODD Measures | Normal (N = 40) | | ODD (N=40) | |
| | M | SD | M | SD |
| CTRS-R:S | | | | |
| OP | 0.07 | 0.35 | 10.00[e] | 3.15 |
| DBD | | | | |
| OP | 0.42 | 0.84 | 15.75[f] | 5.12 |

Note. CTRS-R:S = Conners' Teacher Rating Scale - Revised:Short; Hyp = Hyperactivity Factor; ADHD = Attention-Deficit Hyperactivity Factor; OP - Oppositional Factor. DBD = Disruptive Behavior Disorder Rating Scale; IO = Impulsivity/Overactivity Factor; IN = Inattention Factor; OP = Oppositional Factor.

a, b, c, d ADHD tape > Normal tape, $p < .000$.

e, f ODD tape > Normal tape, $p < .000$.

Hyperactivity Factor, the ratings for the ADHD tapes (M = 12.67, SD = 3.94) were significantly higher than the ratings for the normal tapes (M = 0.62, SD = 0.93, t (39) = 19.09, p < .001). On the Conners' ADHD Index, the ratings for the ADHD tapes (M = 25.82, SD = 6.22) were significantly higher than the ratings for the control tapes (M = 2.22, SD = 2.95, t (39) = 22.25, p < .001). On the Impulsivity/Overactivity Factor from the Disruptive Behavior Disorders Rating Scale (DBD), the ratings for the ADHD tapes (M = 15.82, SD = 5.48) were significantly higher than the ratings for the control tapes (M = 0.67, SD = 1.46, t (39) = 17.22, p < .001) . On the Inattention Factor from the DBD, the ADHD tapes (M = 16.15, SD = 5.15) were rated significantly higher than the control tapes (M = 1.12, SD = 2.29, t (39) = 17.48, p < .001).

This same pattern also held true for the two measures of Oppositional Defiance. On the Conners' Oppositional Factor, the ODD tapes (M = 10.00, SD = 3.15) were rated significantly higher than the control tapes (M = 0.07, SD = 0.35, t (39) = 20.14, p < .001). On the Oppositional Factor from the DBD, the ODD tapes (M = 15.75, SD = 5.12) were rated significantly higher than the control tapes (M = 0.42, SD = 0.84, t (39) = 18.63, p < .001).

ADHD Ratings: ADHD Vs. ODD Tapes.

To test for the presence of the unidirectional negative halo effect, and the influence of child gender on the effect, 2 (type of tape: ADHD vs. ODD) x 2 (target child gender: male vs. female) ANOVAs were conducted for each of the four CTRS-R:S factors and each of the three DBD factors (see Table 5).

Collapsed across child gender, the mean Conners' Hyperactivity Factor score for the ADHD tapes (M = 12.68, SD = 3.94) was significantly higher than for the ODD tapes (M = 5.08, SD = 2.92, F (1, 76) =110.17, p < .001). The main

effect of child gender was not significant, $F$ (1, 76) = .17, $p$ = .68. However, there was a significant interaction between the target-child gender and type of tape, $F$ (1, 76), = 5.19, $p$ = .026. A Tukey's post-hoc analysis indicated that on the ODD tapes, the male child ($M$ = 6.05, $SD$ = 2.56) was rated significantly higher than the female child ($M$ = 4.10, $SD$ = 2.10) $p < .05$) on the Conner's Hyperactivity Factor.

Similar results were found with the Conners' ADHD Index, where, collapsed across gender, the mean score for the ADHD tapes ($M$ = 25.83, $SD$ = 6.22) was significantly higher than the mean score for the ODD tapes ($M$ = 11.90, $SD$ = 6.58), $F$ (1, 76) = 99.99, $p < .001$. Once again, the impact of child gender was not significant, $F$ (1, 76) = .91, $p$ = .344, but there was a significant gender x type of tape interaction, $F$ (1, 76) = 5.53, $p$ = .021. Tukey's post-hoc analysis indicated that the male child on the ODD tape ($M$ = 14.20, $SD$ = 7.08) was rated significantly higher than the female child on the ODD tape ($M$ = 9.60, $SD$ = 5.25), $p < .05$) on the Conner's ADHD Index.

On the DBD scale, similar main effects were found, but the interactions were either not significant or lost their statistical significance upon post-hoc analyses. On the Impulsivity/Overactivity Factor from the DBD, collapsed across gender, the ADHD tapes ($M$ = 15.82, $SD$ = 5.48) were rated significantly higher than the ODD tapes ($M$ = 7.35, 4.49), $F$ (1, 76) = 59.73, $p < .001$. The main effect of gender was not significant, $F$ (1, 76) = .27, $p > .05$. Though the interaction was significant, $F$ (1, 76) = 5.09, $p$ = .027, the differences between the groups did not meet the significance level of $p < .05$ on a Tukey's post-hoc test.

Table 5. Teachers' Mean CTRS-R:S and DBD Ratings for Male and Female Children on the ADHD and ODD Tapes.

| | | CTRS-R:S | | | | | | DBD | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | HYP | | ADHD | | OP | | IO | | IN | | OP | |
| Type of Tape | n | M | SD | M | SD | M | SD | M | SD | M | SD | M | SD |
| ADHD | | | | | | | | | | | | | |
| Male | 20 | 12.00 | 4.30 | 24.85 | 6.55 | 2.80 | 2.78 | 14.30 | 4.77 | 15.90 | 5.41 | 6.30 | 4.64 |
| Fem. | 20 | 13.35 | 3.53 | 26.80 | 5.88 | 5.90 | 4.21 | 17.35 | 5.82 | 16.40 | 4.99 | 13.50 | 9.10 |
| Comb[a] | 40 | 12.68[b] | 3.94 | 25.83[c] | 6.22 | 4.38 | 3.87 | 15.82[d] | 5.48 | 16.15[e] | 5.15 | 9.90 | 8.00 |
| ODD | | | | | | | | | | | | | |
| Male | 20 | 6.05[f] | 2.56 | 14.20[g] | 7.08 | 8.95 | 2.89 | 8.30 | 5.21 | 8.00 | 5.62 | 14.70 | 5.35 |
| Fem. | 20 | 4.10 | 2.10 | 9.60 | 5.25 | 11.00 | 3.12 | 6.40 | 3.52 | 5.60 | 3.82 | 16.80 | 4.70 |
| Comb[h] | 40 | 5.08 | 2.52 | 11.90 | 6.58 | 10.00[i] | 3.15 | 7.35 | 4.49 | 6.80 | 4.89 | 15.75[j] | 5.12 |
| Combined ADHD/ODD[k] | | | | | | | | | | | | | |
| Male | 40 | 9.03 | 4.62 | 19.53 | 8.63 | 5.88 | 4.19 | 11.30 | 5.79 | 11.95 | 6.76 | 10.50 | 6.52 |
| Female | 40 | 8.73 | 5.49 | 18.20 | 10.30 | 8.50[l] | 4.48 | 11.88 | 7.30 | 11.00 | 7.01 | 15.15[m] | 7.37 |

Note. CTRS-R:S = Conners' Teacher Rating Scale - Revised: Short; Hyp = Hyperactivity Factor; ADHD = Attention-Deficit Hyperactivity Factor; OP - Oppositional Factor. DBD = Disruptive Behavior Disorder Rating Scale; IO = Impulsivity/Overactivity Factor; IN = Inattention Factor; OP = Oppositional Factor.

[a]Comb. = ADHD tapes collapsed across gender. [b, c, d,e] Collapsed across gender, ADHD tape > ODD tape, $p < .000$. [f, g] Male ODD tape > Female ODD tape, $p < .05$. [h]Comb. = ODD tapes collapsed across gender. [i, j] Collapsed across gender, ODD tape > ADHD tape, $p < .001$. [k] Combined ADHD/ODD = male and female tapes collapsed across type of tape. [l, m] Collapsed across type of tape, Females > Males, $p < .05$.

On the Inattention Factor from the DBD, collapsed across gender, the ADHD tapes (M = 16.15, SD = 5.15) were rated significantly higher than the ODD tapes (M = 6.80, SD = 4.89), F (1, 76) = 69.67, p < .001. The main effect of gender was not significant, F = .72, p > .05, nor was the gender x type of tape interaction significant, F (1, 76) = 1.68, p = .20.

ODD Ratings: ADHD vs. ODD Tapes

On the Conners' Oppositional Factor, the ODD tapes, collapsed across gender, were rated significantly higher than the ADHD tapes (M = 10.00, SD = 3.15 vs. M = 4.38, SD = 3.87), F (1, 76) = 58.08, p < .001. Collapsed across type of tape, there was also a main effect for the gender of target child, with females (M = 8.50, SD = 4.48) scoring higher than males (M = 5.88, SD = 4.19), F (1, 76) = 12.65, p < .001. The interaction between child gender and type of tape was not significant, F (1, 76) = .51, p = .48.

The DBD Oppositional Factor followed the same pattern as the Conners' Oppositional Factor. Collapsed across gender, the ODD tapes were rated significantly higher on the Oppositional Factor than the ADHD tapes (M = 15.75, SD = 5.12 vs. M = 9.90, SD = 8.00, F (1, 76) = 17.56, p < 001. There was also a significant gender main effect, collapsed across type of tape, in which females (M = 15.15, SD = 7.37) scored higher than males (M = 10.50, SD = 6.52), F (1, 76) = 11.09, p = .001. The interaction effect was not significant, F (1, 76) = 3.34, p > .05.

DSM-III (R) Ratings.

Since the DBD rating scale contains the 23 DSM-III (R) diagnostic criteria for ADHD and ODD, worded as closely as possible to the actual DSM-III (R), the scale was used to tally the number of DSM-III (R) criteria endorsed by each

teacher for each tape rated. To determine the proportion of teachers who rated the tapes as exhibiting the number of symptoms required for DSM-III (R) diagnostic cutoffs (eight symptoms for ADHD and five symptoms for ODD) frequency counts of the tapes meeting criteria were computed. Since the DBD employs a 4-point rating scale (Not true at all = 0; Just a little true = 1; Pretty much true = 2; and Very much true = 3) a cutoff-point had to be determined regarding at what level (0-3) an endorsement would be considered as meeting the DSM-III-R threshold for the presence of the symptom. The same criteria used by Abikoff, Courtney, Pelham, and Koplewicz (1993) with the similar SNAP-III rating scale was employed; an item had to be endorsed as occurring at least "pretty much" before it was counted.

As indicated in Table 6, the ratings of 50% of the teachers classified the children in the ADHD tapes as meeting DSM-III (R) criteria for ADHD, whereas the tapes depicting the ODD children were only rated as meeting ADHD criteria by 12.5% of the teachers, a significant difference, $\chi^2$ (1, $\underline{N}$ = 80) = 12.56, $\underline{p} < .01$. A similar pattern held for ODD ratings, where 62.5% of the teachers classified the child in the ODD tapes as meeting DSM-III (R) criteria for ODD, while only 7.5% of the teachers classified the child in the ADHD tapes as meeting the criteria for ODD, $\chi^2$ (1, $\underline{N}$ = 80) = 17.29, $\underline{p} < .01$. Significant gender differences were found in the ODD classifications of the ADHD tapes, where more females than males were classified as ODD, $\chi^2$ (1, $\underline{n}$ = 40) = 4, $\underline{p} < .05$. Also, more females than males were classified as having both ADHD and ODD on the ADHD tape, $\chi^2$ (1, $\underline{n}$ = 40) = 4, $\underline{p} < .05$.

Table 6. Percentage of Teachers Whose Ratings Met DSM-III (R) Diagnostic Criteria for ADHD, ODD, and Combined ADHD and ODD

| | Percentage whose ratings met diagnostic criteria for: | | |
|---|---|---|---|
| Type of Tape | ADHD | ODD | ADHD & ODD |
| ADHD | | | |
| Male (n = 20) | 50.0 | 0.0 | 0.0 |
| Female (n = 20) | 50.0 | 25.0[a] | 20.0[b] |
| Combined[c] (n = 40) | 50.0[d] | 12.5 | 10.0 |
| ODD | | | |
| Male (n = 20) | 10.0 | 50.0 | 10.0 |
| Female (n = 20) | 5.0 | 75.0 | 0.0 |
| Combined[e] (n = 40) | 7.5 | 62.5[f] | 5.0 |

Note: ADHD = Attention-Deficit Hyperactivity Disorder; ODD = Oppositional Defiant Disorder.

[a, b]ADHD female tape > ADHD male tape, $p < .05$.

[c]Combined = males and females from the ADHD tapes combined.

[d]ADHD tape > ODD tape, $p < .05$.

[e]Combined = males and females from the ODD tapes combined.

[f]ODD tape > ADHD tape, $p < .05$.

## Secondary Analyses

Since the ODD and normal tapes did not significantly differ on the mean number of hyperactive behaviors recorded on the Revised Stony Brook Observation Code (see Table 7), $t(4) = 2.62$, $p > .05$; nor on the mean number of 15-second intervals during which ADHD behaviors occurred on the COCADD

(see Table 2), $\chi^2$ (1, N = 140) = .47, p > .05; analyses were conducted to see how the two tapes compared on the teacher-based ADHD ratings.

Comparison of the ADHD and normal tapes on the oppositional defiant ratings obtained from the Revised Stony Brook Observation Code revealed similar results: the two tapes did not significantly differ on the mean number of ODD behaviors recorded (see Table 8), t (4) = 1.00, p > .05; nor on the percentage of 15-second intervals during which ODD behaviors were recorded on the COCADD (see Table 2), $\chi^2$ (1, N = 140) = .72, p > .05; hence, analyses were undertaken to see how the ADHD and normal tapes compared on the teacher-based ODD ratings.

In addition, the above mentioned analyses were designed to evaluate the impact of the target child's gender. Thus, mixed statistical designs were employed, using teacher ratings on the pathology and normal tapes as the repeated measure, and the target child's gender as the between-subjects factor. The dependent measures used were: the Conners' Hyperactivity Index and ADHD Factor, and the Impulsivity/Overactivity and Inattention Factors from the DBD. The comparisons between the ADHD and normal tapes were conducted on the Conners' Oppositional Factor and the DBD's ODD Factor. All reported F statistics come from Hotelling's Trace multivariate tests.

ADHD Ratings: ODD vs. Normal Tapes

On the Conners' Hyperactivity Factor, teacher ratings of the ODD tapes were significantly higher than teacher ratings on the normal tapes (M = 5.27, SD = 2.51 vs. M = 1.20, SD = 1.36, F (1, 38) = 114.61, p < .001). Though the main effect of child gender was not significant, F (1, 38) = 2.54, p = .12, there was a

Table 7. Trained Observers Mean Hyperactivity-Composite and Oppositional Defiant-Composite Ratings on the Revised Stony Brook Observation Code for ODD and Normal Tapes.

| | Type of Tape | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ODD | | | | | | Normal | | | | | |
| | Male (N=5) | | Female (N=5) | | Total (N=10) | | Male (N=5) | | Female (N=5) | | Total (N=10) | |
| Scoring System | M | SD | M | SD | M | SD | M | SD | M | SD | M | SD |
| HYP | 14.80 | 1.48 | 14.80 | 1.10 | 14.80 | 1.29 | 13.00 | 2.35 | 12.80 | 1.48 | 12.90 | 1.91 |
| OP | 8.00 | 0.71 | 7.60 | 1.14 | 7.80[a] | 0.92 | 0.20 | 0.50 | 1.00 | 0.00 | 0.60 | 0.25 |

Note. Stony Brook = Revised Stony Brook Observation Code: HYP = hyperactivity composite score; OP = oppositional defiant composite score.

[a]ODD tape > Normal tape, $p < .05$.

Table 8. Trained Observers Mean Hyperactivity-Composite, and Oppositional Defiant-Composite Ratings on the Revised Stony Brook Observation Code, for ADHD and Normal Tapes.

| | Type of Tape | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ADHD | | | | | | Normal | | | | | |
| | Male (N=5) | | Female (N=5) | | Total (N=10) | | Male (N=5) | | Female (N=5) | | Total (N=10) | |
| Scoring System | M | SD | M | SD | M | SD | M | SD | M | SD | M | SD |
| HYP | 50.20 | 5.26 | 52.20 | 7.01 | 51.20[a] | 6.13 | 13.00 | 2.35 | 12.80 | 1.48 | 12.90 | 1.91 |
| OP | 0.80 | 0.84 | 1.20 | 0.45 | 1.00 | 0.64 | 0.20 | 0.45 | 1.00 | 0.00 | 0.60 | 0.22 |

Note. Stony Brook = Revised Stony Brook Observation Code: HYP = hyperactivity composite score; OP = oppositional defiant composite score.

[a]ADHD tape > Normal tape, $p < .05$.

significant interaction, $F$ (1, 38) = 10.54, $p$ = .002 . A Tukey's post-hoc test indicated that the males on the ODD tapes were rated significantly higher than the females on the ODD tapes ($M$ = 6.05, $SD$ = 2.56 vs. $M$ = 4.10, $SD$ = 2.10) (see Table 8).

Findings from the Conners' ADHD Factor indicated that the teachers rated the ODD tapes significantly higher than they rated the normal tapes ($M$ = 11.90, $SD$ = 6.58 vs. $M$ = 3.70, $SD$ = 3.63, $F$ (1, 38) = 114.61, $p$ < .001). Though the main effect of child gender was not significant, $F$ (1, 38) = 2.54, $p$ = .12, there was a significant interaction, $F$ (1, 38) = 11.97, $p$ = .001, with a Tukey's post-hoc test finding that the males on the ODD tape ($M$ = 14.20, $SD$ = 7.08) were rated significantly higher than the females on the ODD tape ($M$ = 9.60, $SD$ = 5.25) (see Table 7).

A significant difference was also found on the DBD Impulsivity/Overactivity Factor, with the ODD tape ($M$ = 7.35, $SD$ = 4.49) rated significantly higher than the normal tape ($M$ = 1.62, $SD$ = 2.11, $F$ (1, 38) = 114.85, $p$ < .001). The tape x gender interaction was not significant, $F$ (1, 38) = 3.00, $p$ = .09), nor was the main effect of gender significant, $F$ (1, 38) = 1.02, $p$ = .32 (see Table 7).

On the final measure of ADHD, the DBD Inattention Factor, teachers rated the ODD tapes significantly higher than the normal tapes ($M$ = 6.80, $SD$ = 4.89 vs. $M$ = 1.75, $SD$ = 2.37, $F$ (1, 38) = 56.22, $p$ < .001. The main effect of child gender was not significant, $F$ (1, 38) = 1.01, $p$ = .32; however, the tape x gender interaction was significant, $F$ (1, 38) = 4.32, $p$ = .04. A Tukey's post-hoc test revealed that the males on the ODD tape ($M$ = 8.00, $SD$ = 5.62) were rated significantly higher than the females on the ODD tape ($M$ = 5.60, $SD$ = 3.81).

ODD Ratings: ADHD vs. Normal Tapes

On the Conners' Oppositional Factor, teacher ratings of the ADHD tapes were significantly higher than teacher ratings on the normal tapes ($M = 4.37$, $SD = 3.87$ vs. $M = 0.08$, $SD = 0.35$, $F(1, 38) = 60.05$, $p < .001$). The main effect of child gender was also significant, $F(1, 38) = 7.16$, $p = .01$, with the females ($M = 3.00$, $SD = 2.23$) being rated higher than the males ($M = 1.95$, $SD = 1.62$). The tape x child interaction, $F(1, 38) = 8.31$, $p = .002$ was also significant, with a Tukey's post-hoc test finding that the females on the ADHD tape ($M = 5.95$, $SD = 4.21$) were rated significantly higher than the males on the ADHD tape ($M = 2.80$, $SD = 22.78$) (see Table 9).

A significant difference was also found on the DBD Oppositional Factor, where teachers rated the ADHD tapes significantly higher than the normal tapes ($M = 9.90$, $SD = 8.00$ vs. $M = 0.17$, $SD = 0.50$, $F(1, 38) = 72.2$, $p < .001$). The main effect of gender was significant, $F(1, 38) = 10.58$, $p = .002$) with teachers rating females ($M = 6.90$ , $SD = 4.88$) higher than males ($M = 3.18$, $SD = 2.41$ ). The tape x gender interaction was also significant, $F(1, 38) = 9.21$, $p = .004$, with the Tukey's post-hoc test indicating that the females on the ADHD tapes ($M = 13.50$, $SD = 9.10$) were rated significantly higher than the males on the ADHD tapes ($M = 6.30$, $SD = 4.46$) (See Table 10).

Table 9. Teacher Mean ADHD Ratings on the CTRS-R:S and DBD Rating Scales for the ODD and Normal Tapes.

| | Type of Tape | | | | | | | | | | | | Effect | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ODD | | | | | | Normal | | | | | | Size Indices[a] | |
| | Male (n = 20) | | Female (n = 20) | | Total (n = 40) | | Male (n = 20) | | Female (n = 20) | | Total (n = 40) | | | |
| Factor Scores | M | SD | M | SD | M | SD | M | SD | M | SD | M | SD | Tape | Gender |
| CTRS-R:S | | | | | | | | | | | | | | |
| HYP | 6.05[b] | 2.56 | 4.10 | 2.10 | 5.07[c] | 2.51 | 1.00 | 1.21 | 1.40 | 1.50 | 1.20 | 1.36 | 0.75 | 0.22 |
| ADHD | 14.20[d] | 7.08 | 9.60 | 5.25 | 11.90[e] | 6.58 | 3.25 | 3.61 | 4.15 | 3.69 | 3.70 | 3.63 | 0.74 | 0.24 |
| DBD | | | | | | | | | | | | | | |
| IO | 8.30 | 5.21 | 6.40 | 3.52 | 7.35[f] | 4.49 | 1.65 | 2.03 | 1.60 | 2.23 | 1.62 | 2.11 | 0.75 | 0.07 |
| IN | 8.00[g] | 5.62 | 5.60 | 3.81 | 6.80[h] | 4.89 | 1.55 | 2.33 | 1.95 | 2.46 | 1.75 | 2.37 | 0.60 | 0.10 |

Note. CTRS-R :S = Connors' Teacher Rating Scale - Revised :Short; Hyp = Hyperactivity Factor; ADHD = Attention-Deficit Hyperactivity Index. DBD = Disruptive Behavior Disorder Rating Scale; IO = Impulsivity/Overactivity Factor; IN = Inattention Factor.

[a] Effect Size Index = Difference in ratings between the ADHD and the normal tapes, and difference in ratings between genders. Effect sizes: $\eta^2$ = .01 is small, $\eta^2$ = .06 is medium, $\eta^2$ = .14 is large.

b, d, g Male ODD tape > female ODD tape, $p < .05$.

c, e, f, h ODD tape (male & female combined) > Normal tape (male & female combined), $p < .05$.

Table 10. Teacher Mean ODD Ratings on the CTRS-R:S and DBD Rating Scales for the ADHD and Normal Tapes.

| | Type of Tape | | | | | | | | | | | | Effect | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ADHD | | | | | | Normal | | | | | | Size Indices[a] | |
| | Male (n = 20) | | Female (n = 20) | | Total (n = 40) | | Male (n = 20) | | Female (n = 20) | | Total (n = 40) | | | |
| Factor Scores | M | SD | M | SD | M | SD | M | SD | M | SD | M | SD | Tape | Gender |
| CTRS-R:S | | | | | | | | | | | | | | |
| OP | 2.80 | 2.78 | 5.95[b] | 4.21 | 4.37[c] | 3.87 | 0.10 | .45 | 0.05 | 0.22 | .08 | 0.35 | 0.61 | 0.18 |
| DBD | | | | | | | | | | | | | | |
| OP | 6.30 | 4.64 | 13.50[d] | 9.10 | 9.90[e] | 8.00 | 0.05 | .22 | 0.30 | 0.65 | 0.17 | 0.50 | 0.65 | 0.20 |

Note. CTRS-R :S = Connors' Teacher Rating Scale - Revised (Short); OP - Oppositional Factor. DBD = Disruptive Behavior Disorder Rating Scale; OP = Oppositional Factor.

[a] Effect Size = Difference in ratings between the ADHD and normal tapes, and difference in ratings between genders. Effect size: $\eta 2 = .01$ is small, $\eta 2 = .06$ is medium, $\eta 2 = .14$ is large.

[c, e] ADHD tapes (male & female combined) > Normal tapes (male & female combined), $p < .000$.

[b, d] Female ADHD tape > Male ADHD tape, $p < .05$.

# CHAPTER IV

# DISCUSSION

Since the present study is a partial replication of two earlier studies (Abikoff, et al., 1993; Stevens et al., 1998) discussion of the results necessitates placing it within the context of these earlier studies.

Abikoff, et al. (1993) used an analogue research design in which elementary school teachers viewed tapes of a fourth grade classroom. In each tape one target child portrayed behaviors typical of a normal child, or behaviors symptomatic of a child with either ADHD or ODD. The authors found evidence of a unidirectional bias in teachers' ratings (negative halo effect). Teachers accurately rated hyperactivity and other ADHD symptomatic behaviors when the student exhibited ADHD-type behaviors; however, when the student engaged in ODD-type behaviors, the teachers' ratings of hyperactivity and other ADHD-type behaviors were spuriously inflated. In contrast, the teachers accurately rated ODD symptomatic behaviors regardless of the presence of hyperactivity and other ADHD-type behaviors.

Using the same tapes made by Abikoff et al. (1993), Stevens et al. (1998) partially replicated Abikoff et al.'s results. Stevens et al. found that teachers did not rate the ODD and ADHD tapes significantly different on the Inattention/Passive Factor, or the Hyperkinesis Index from the CTRS-28, thus providing evidence for the presence of the negative halo effect. However, the ADHD and ODD tapes were rated significantly different on three measures: the Hyperactivity Factor from the CTRS-28, and on both ADHD measures from the

SNAP-IV Rating Scale. Though Stevens et al. had found statistically reliable differences between the ADHD and ODD tapes on these three ADHD measures, they noted that the magnitude of the main effect sizes were small ($\omega^2$ = .27, .50 and .38, respectively) in comparison to the effect sizes found between the two tapes on the two ODD measures: the CTRS-28 Conduct Problems Factor ($\omega^2$ = 1.23) and the SNAP-IV ODD Scale ($\omega^2$ = 1.08). The authors concluded that the oppositional defiant behaviors had indeed exerted a unidirectional negative halo effect on teacher ratings of ADHD.

The present study used different tapes which followed the same scripts, portrayed the same disruptive behaviors, and ran the same length of time, yet did not replicate the findings of Abikoff et al. (1993). The ADHD and ODD tapes in the present study were rated by teachers as being significantly different on all four ADHD measures and on the two ODD measures. These findings are partially consistent with the results from Stevens et al. (1998) who found significant differences between the ADHD and ODD tapes on three of five ADHD measures and both ODD measures. However, in contrast to Stevens et al. who found large effect sizes between the two tapes on the ODD measures and only moderate effect sizes on the ADHD measures, the present study found large effect sizes on the ODD measures ($\eta^2$ = .19 and .43) and still larger effect sizes on the ADHD measures ($\eta^2$ = .48, .54, .56, and .59). These effect sizes suggested that the presence of ADHD behaviors may have exerted more of a biasing effect on teacher ratings of oppositionality, than the presence of ODD behaviors had exerted on teacher ratings of ADHD.

One explanation for this near reversal in the negative halo effect between Abikoff et al. (1993) and the present study, may have begun with the findings from Stevens et al. (1998), as the former study had provided evidence for a stronger halo effect than the latter had found. Stevens et al. concluded that a possible explanation for their discrepant results was an increase in teacher awareness regarding the different types of childhood behavior disorders during the eight years between the data collection for the two studies. The authors noted that a shift in educational philosophy during the 1990's had increased mainstreaming and exposed more teachers to children with ADHD, which may have decreased the strength of the halo effect. This explanation, however, seems unlikely given that Stevens et al. found empirical evidence that greater knowledge and experience with ADHD was not associated with more accurate teacher ratings.

A second possible explanation for the attenuated halo effect found by Stevens et al. (1998), and the further reduction found in the present study, may lie in the different locations from which teachers were drawn. The teachers in Abikoff et al. (1993) were drawn from the New York public school system, and the teachers in Stevens et al. were recruited from a medium-sized Midwestern city, while the teachers from the present study were drawn primarily from schools spread throughout rural Northeastern North Dakota and Northwestern Minnesota, including one small Midwestern city. Between Abikoff et. al and the present study there was a shift from east to Midwest, and from urban to rural.

Pekkansen (2000) provided evidence which may indirectly corroborate this theory. Pekkansen found large variations in Ritalin prescriptions, and consequently ADHD diagnoses, from one part of the United States to another.

Dr. Peter Jensen, from the National Institute of Mental Health, stated that the national ADHD diagnostic pattern resembles a "patchwork quilt", because children "are both over-diagnosed and under-diagnosed depending on where they live and the attitudes within their communities" (Pekkarisen, 2000, p. 157). A 1991-92 survey found that children in some counties of New York state are 10 times more likely to be prescribed Ritalin than are children from other counties from New York (Pekkansen, 2000). In a second study, Offord et al. (1987) compared the six-month prevalence rates of hyperactivity, and rates of mental health service utilization for hyperactivity, between urban and rural areas of Ontario. The study found a significantly higher rate of hyperactivity in urban compared to rural areas ($\underline{M}$ = 7.0 vs. 4.6, diagnoses per 100 children). The study also found that urban and rural children did not significantly differ on rates of utilization of mental health services.

In contrast, Daley, Onwuegbuzie, and Griffin (1998) found no correlation between the size of school districts and ADHD prevalence rates in a Mid-southern state. Generalizations made from these two studies may be tenuous, as the results from the Canadian study may not pertain to the United States, while the results from Daley et al. suffered from a very low return rate (41%) on the surveys sent to school superintendents. However, if ADHD prevalence patterns do fluctuate widely depending upon geographic location within the country, then it is equally likely that teachers' perceptions of what constitutes ADHD-type behaviors would also fluctuate. Accordingly, such fluctuations in perceptions would appear on teacher-completed rating scales.

If Stevens et al.'s (1998) conclusion that increased teacher awareness of the behavioral disorders leads to decreased bias on ADHD rating scales is

valid, and the assumption that this knowledge base has increased over the past decade is accurate, then it stands to reason that the present study, which collected data five years after Stevens et al., might find the halo effect even further diluted. In addition, if rural Midwestern teachers are less likely to perceive all disruptive behavior as being ADHD in nature, compared to urban Eastern teachers, then the halo effect would be even further weakened in the present study. Yet, the dissemination of information regarding childhood behavior disorders, increased teacher exposure to behavior disordered children, and geographic location within the country, seem wholly inadequate to explain the near reversal in the halo effect found in the present study. A more plausible explanation may lie in the differences between the tapes made by Abikoff et al. (1993) and the tapes utilized in the present study.

## Tape Evaluation

Comparison of the validation results from the Abikoff et al. (1993) tapes and the tapes used in the present study reveal several important findings. Based on the Revised Stony Brook Observation Code, the target child in the ODD tape from Abikoff et al. exhibited 40% more ADHD behaviors than the target children on the ODD tapes from the present study. In addition, based on the COCADD results, the target child in Abikoff et al.'s ODD tape exhibited ADHD behaviors during 11% more of the 15-second intervals, compared to the target children on the ODD tapes from the present study. Taken together, these findings suggest that the target child in the ODD tapes from Abikoff et al. engaged in more ADHD behaviors, and these behaviors were more evenly distributed throughout the duration of the tape, compared to the target children on the ODD tapes in the present study.

Comparison of the Revised Stony Brook Hyperactivity Composite scores from the ADHD tape from Abikoff et al. (1993) and the ADHD tapes from present study revealed considerable similarity (M = 51 vs. 54). However, the ADHD Composite scores from the COCADD revealed that the children on the ADHD tapes in the present study engaged in ADHD behaviors during 19% more of the 15-second intervals than did the child on the ADHD tape from Abikoff et al.

The direct comparison of factor scores derived from the teacher ratings for the different tapes in Stevens et al. (1998), Abikoff et al. (1993) and the present study could help clarify why the unidirectional negative halo effect was not found in the present study. For example, if Abikoff et al.'s ODD tapes did contain considerably more ADHD behaviors than the ODD tapes in the present study, then Abikoff et al.'s ODD tapes would be expected to score higher on ADHD rating scales compared to the ODD tapes from the present study.

Unfortunately, methodological differences among the three studies do not allow for valid comparisons between the factor scores obtained by each study. First, Abikoff et al. (1993) used the CTRS-28 and direct DSM-III-R descriptors, while Stevens et al. (1998) used the CTRS-28 and SNAP-IV Rating Scale, and the present study utilized the CTRS-R (S) and the DBD Rating Scale. Because each study used different measures, the direct comparison of factor scores is impossible. In addition, the two earlier studies used only male target children, while the present study utilized both males and females. The Conners and SNAP scales employ gender-specific norming distributions in the transformation of raw scores into scaled scores. Since the earlier studies only reported scaled scores, raw scores from the current study would need to be transformed into scaled scores before comparisons between the three studies

could be made. Methodologically, this was problematic, as the current study needed to determine how to transform the raw scores obtained from the teachers who viewed the female tapes. If the female norms were used in the transformation, scores from the female tapes would have been inflated relative to the scores from the male tapes; yet, the use of male norms for the transformation of female scores was not justifiable. As a result of these two major methodological differences, the current study was not a valid replication of Abikoff et al., nor Stevens et al., and factor scores could not be readily compared.

## Explanation of Tape Differences

The tapes from Abikoff et al. (1993) and the present study followed the same scripts and same blocking, contained the same number of children in the classes and had classrooms of equal geographic size. The tapes were the same length and the audiovisual qualities were comparable. Consequently, it is difficult to account for the considerable differences detected by trained observers. One explanation for the differences may lie with the strict timing requirements used by the Revised Stony Brook Observation Code. Several of the behavior categories, such as "off task" and "non-compliant", require that a codeable behavior begin during one 15-second interval and continue throughout the next 15-second interval before it is recorded. Thus, a behavior may occur for as long as 28 seconds and not be recorded, while another behavior may occur for as few as 16 seconds and be counted. During the creation of the tapes for the present study, no attempt was made to replicate the exact duration of each instance of disruptive behavior, other than those few behaviors where exact durations were specifically noted in the Abikoff et al.

script. As a result, fewer of the ADHD-type behaviors exhibited by the ODD children in the present study met the timing requirements to be recorded. This would have attenuated the Revised Stony Brook ADHD scores for the ODD tapes. This also helps partially explain why teachers in the present study rated the ODD tapes considerably lower on ADHD measures compared to teacher ratings of the ODD tapes made by Abikoff et al.; many instances of ADHD-type behavior may have been shorter in duration on the tapes from the present study, thus not reaching the "critical level" necessary to be noted as disruptive by the teachers. Indeed, many teachers seemed to hold this subjective opinion, as they frequently made the following type of comment after viewing the ODD tapes: "That kid was mild compared to . . . from my class".

Differences between the Abikoff et al. (1993) tapes and the tapes from the present study could also be the result of natural, preexisting differences between the child actors. Regardless of similar lines and behaviors, some actors are simply more convincing than others. In addition, the tapes had different directors who, no doubt, elicited different tones of voice and voice inflections, as well as different behavioral pronouncements from the children. Indeed, some kinesic researchers, such as Mehrabian (1971), have suggested that nonverbal communication sent by gestures, facial expressions, eye contact, postural variation, and touch may account for up to 90% of all message transmission .

ADHD Ratings: ODD vs. Normal Tapes

As mentioned at the end of the Introduction section of this paper, the pathology and normal tapes would be compared if the results of the validation process indicated that such comparisons were warranted. Given the validation

results from the Revised Stony Brook Observation Code and the COCADD, it was deemed that the more appropriate way to evaluate for the biasing effect of oppositional defiant behaviors on teacher ratings of ADHD was to compare the ODD and normal tapes on ADHD measures. In keeping with methodological requirements, the two tapes significantly differed on the total number of halo-suspected ODD behaviors exhibited (M = 7.80 vs. 0.50), and on the percentage of 15-second intervals during which these ODD behaviors occurred (21.5% vs. 0.0%). Conversely, the ODD and normal tapes did not differ significantly on the total number of ADHD behaviors exhibited (M = 14.80 vs. 12.90), nor on the percentage of 15-second intervals that contained ADHD behaviors (18.0% vs. 15.0%). Consequently, any differences found on teacher-based ADHD measures between these two groups could be attributed to the biasing effect of oppositional defiant behaviors on teacher-based ADHD ratings.

When teacher ratings for the ODD and normal tapes were compared on each of the four ADHD scales, significant evidence was found that the presence of oppositional defiant behaviors had inflated teacher ratings of ADHD-type behaviors (see Table 9). In addition, the hypothesis that this bias would be less pronounced when using the more operationally defined DBD Rating Scale, as opposed to the more global Conners' Scale, was not supported, as may be seen by comparing the effect sizes found with the two measures. Cohen (1977) characterized effect sizes with the following categories: $\eta^2 = .01$ is small, $\eta^2 = .06$ is medium, and $\eta^2 = .14$ is large. Accordingly, the effect sizes found with the CTRS-R:S Hyperactivity Factor ( $\eta^2 = .75$) and the ADHD Index ($\eta^2 = .74$) were both extremely large. Likewise, the effect size found with the DBD

Impulsivity/Overactivity Factor was large ($\eta^2$ = .75) as was the effect size found with the DBD Inattention Factor ($\eta^2$ = .60). One possible reason that more substantial differences were not found between the Conners' Rating Scale and the DBD is that the present study used the recently revised Conners' Teacher Rating Scale (CTRS-R:S), as opposed to the CTRS-28. Several of the more global statements found in the CTRS-28, which had been used by Abikoff et al. (1993) and Stevens et al. (1998), such as, "Pouts and sulks", "Acts smart", and "Childish and immature", have been replaced with more descriptive items to bring it in line with the language of the DSM-IV (Conners, 1997).

Impact of Child Gender

Evaluation of the Revised Stony Brook and COCADD validation results for the male and female ODD tapes in the present study indicated that the two tapes met the requirements necessary for comparison purposes. The male and female tapes did not significantly differ on the Oppositional Composite score from the Revised Stony Brook Observation Code (M = 8.00 and 7.60 respectively), nor on the percentage of 15-second intervals during which oppositional defiant behaviors occurred, as measured by the COCADD (21.7% vs. 21.4%, respectively). In addition, the male and female ODD tapes exhibited the same number of ADHD-type behaviors, as measured by the Stony Brook Hyperactivity Composite (M =14.80 vs. 14.80, respectively), and they exhibited these ADHD-type behaviors during a similar number of 15-second intervals, as measured by the COCADD (14.3% vs. 15.7%, respectively). Thus, if the male and female tapes, which exhibited similar levels of ODD and ADHD-type

behaviors, differed on teacher-based ADHD rating scales, this difference could logically be attributed to the impact of the child's gender.

The results of this comparison supported our hypothesis. Males exhibiting ODD behaviors scored higher on teacher-based ADHD rating scales, compared to females exhibiting the same ODD behaviors. The males scored significantly higher on both the Hyperactivity Factor and the ADHD Index from the CTRS-R:S, and on the Inattention Factor from the DBD. These differences were statistically reliable with large size effects and suggested that the unidirectional negative halo effect of oppositional defiant behaviors on teacher-based ADHD ratings had differentially impacted males and females. These results supported the central thesis of this paper: the gender-based prevalence differential found with ADHD may be partly attributable to a large gender bias in the negative halo effect.

Although a stronger bias emerged when teachers rated males, the presence of ODD-type behaviors also exerted a considerable influence on teacher ratings of females. The female ODD tape was rated higher than the female normal tape on all four ADHD measures, though they had not significantly differed on the number of ADHD behaviors exhibited, as measured by the Stony Brook and the COCADD. Thus, it appeared that, regardless of the gender of the child, teachers failed to differentiate between the hyperactive, inattentive, and impulsive behaviors associated with ADHD, and the oppositional and defiant behaviors characteristic of ODD. Ultimately, this lack of behavior differentiation resulted in the artificial inflation of scores on teacher-based ADHD rating scales for both males and females.

It is possible that the artificial nature of this analogue study may have attenuated the actual magnitude of the gender-based difference in teacher ratings. For example, the teachers in the present study viewed only a ten-minute segment of the child's behavior, when ordinarily they would have viewed the child in a variety of settings over a considerably longer period of time. In addition, during the study the child had the teacher's undivided attention, which rarely occurs during the course of the school day.

It is possible that the bias toward rating males exhibiting ODD-type behaviors higher on ADHD rating scales, compared to similarly behaving females, may result from a gender-based stereotype held by many teachers. To stereotype is to generalize. In order to simplify the world, people often generalize, and they are more likely to do so when pressures make simplification more necessary (Myers, 1999). Such may be the case for classroom teachers who face various pressures and time demands. Several researchers have shown that teacher ratings show more strongly pronounced differences between the genders in ADHD behaviors than do parent ratings (Achenbach,1990; Breen & Altepeter, 1990; McGee & Feehan,1991; Szatmari, Offord, & Boyle, 1989). These findings could partially result from the various pressures facing teachers, which could make them more likely to rely on generalizations as compared to parents.

In summary, the short duration of the tapes, the teachers' focused attention, and the lack of any explicitly created pressures on the teachers, could have all contributed to an attenuated gender-based differential found for the negative halo effect in the present study. Accordingly, one recommendation for further research would be to place teachers doing the ratings in a more

pressured environment and see if there is a corresponding increase in the gender-based differences with the negative halo effect. A second recommendation would be to significantly increase the duration of the tapes. Ideally, the tapes should show the child involved in a variety of activities, in different settings, and over a longer period of time. A third recommendation for future research would be to evaluate the relationship between teacher characteristics and the strength of these biases. Abikoff et al. (1993) found that regular education teachers, compared to special education teachers, tended to rate students higher on ADHD rating scales. Conversely, Stevens et al. (1998) found that level of education and knowledge of ADHD were not related to the degree of bias exhibited by teachers. It is possible that these biases may be related to teacher gender; however neither of the two previous studies, nor the present study, have utilized enough male teachers to evaluate this possibility. Furthermore, the relationship between teacher personality characteristics and the strength of these biases could be examined.

### ODD Ratings: ADHD vs. Normal Tapes

Methodologically, the comparison of the ADHD and normal tapes on ODD measures was also warranted. Based on the validation results from the Revised Stony Brook Observation Code and the COCADD, the ADHD and normal tapes differed significantly on the total number of halo-suspected ADHD behaviors exhibited ($\underline{M}$ = 50.60 vs. 12.90), and on the percentage of 15-second intervals during which these ADHD behaviors occurred (69.0% vs. 15.0%). The ADHD and normal tapes did not differ significantly on the total number of ODD behaviors exhibited ($\underline{M}$ =1.00 vs. 0.40), nor in the percentage of 15-second intervals that contained ODD behaviors (0.0% vs. 3.6%). Thus, any differences

found on the teacher-based ODD measures between these two groups could be attributed to the biasing effect of ADHD behaviors on teacher-based ODD ratings.

In contrast to the results from Abikoff et al. (1993) and Stevens et al. (1998) who found the negative halo effect to be unidirectional, the present study found the bias to be bidirectional. When the ADHD and normal tapes were compared on the two ODD rating scales, strong evidence was found that the presence of ADHD-type behaviors had inflated teacher ratings on the CTRS-R:S Oppositional Factor, and on the DBD Oppositional Factor (see Table 8). Though not anticipated, this result was not without precedent. Stevens-Long (1973), who also used videotapes, found a halo effect of overactivity on ratings of aggression. Overactive children were judged to be more aggressive than normally active children, even though the rates of aggressive acts were equal on both tapes. Again, Schachar, Sandberg, and Rutter (1986) found that both "inattention" and "positive interactions with peers" were related to artificially-inflated teacher ratings of behavior problems. The authors concluded that this effect was due to the "perceived nuisance created by children who are overly social or who do not attend to the assigned tasks" (p. 343).

The findings from Schachar et al. (1986) are particularly applicable to the present study. The Hyperactivity Composite score from the Revised Stony Brook Observation Code is composed of six behavior categories. Of these six categories, "off task" and "interference", though representing only 33% of the categories, accounted for 63% of the ADHD male's Hyperactivity composite score, and 51% of the ADHD female's Hyperactivity composite score. Based on the operational definitions of these two categories, "off-task" reflects the

construct of inattention, while "interference" partly reflects positive interactions with peers. Thus, it is possible that the teachers viewed the frequent off-task and interfering behaviors of the ADHD children as creating a nuisance and, accordingly, rated them high on measures of oppositionality and defiance.

Once again, the hypothesis that this bias would be less pronounced when using the more operationally defined DBD Rating Scale was not supported, as may be seen by comparing the size effects found with the two measures ($\eta^2$ = .61 and .65).

Impact of Child Gender

Evaluation of the Revised Stony Brook and COCADD validation results for the male and female ADHD tapes indicated that the two tapes met the requirements necessary to be compared to one another. The male and female tapes did not significantly differ on the Hyperactivity Composite score from the Revised Stony Brook Observation Code (M = 49.00 vs. 52.20, respectively), nor on the COCADD ADHD Composite, which measures the percentage of 15-second intervals during which ADHD-type behaviors occurred (74.0% vs. 65.0%, respectively). In addition, the male and female ADHD tapes did not significantly differ on the expression of ODD-type behaviors as measured by the Stony Brook Hyperactivity composite (M = 0.80 vs. 1.20, respectively), and they portrayed these ODD behaviors during a similar number of 15-second intervals, as measured by the COCADD (2.9% vs. 4.3%, respectively). Thus, if the male and female tapes, which exhibited similar levels of ADHD and ODD-type behaviors, differed on teacher-based ODD rating scales, this difference could logically be attributed to the impact of the child's gender.

Not anticipated in the original hypothesis, the current study found that the presence of ADHD-type behavior created a halo effect on teacher-based ODD ratings, and this effect was more pronounced when females engaged in ADHD-type behaviors, compared to when males exhibited ADHD-type behaviors. The females who exhibited such behaviors scored significantly higher on the Oppositional Factors from both the CTRS-R:S and the DBD, compared to males who displayed similar ADHD-type behaviors. These findings were all the more striking given that the males on the ADHD tape had actually engaged in ADHD-type behaviors during 9% more of the 15-second intervals, compared to the females from the ADHD tapes. These differences were statistically reliable with large size effects and supported the conclusion that the negative halo effect of ADHD behavior on teacher-based ODD ratings had differentially impacted males and females. This finding suggested that teachers may perceive females who exhibit ADHD type behaviors as being more oppositional and defiant than their male counterparts.

Although this bias was stronger when teachers rated females, the presence of ADHD-type behaviors also exerted a considerable influence on the ODD ratings of males. The teachers rated the male ADHD tape higher than the male normal tape on both ODD measures, although the validation results from the Revised Stony Brook and COCADD indicated that the two tapes had not significantly differed on the number of ODD behaviors exhibited. Overall, regardless of the gender of the ADHD child, teachers failed to differentiate between ODD and ADHD-type behaviors, and this failure resulted in the artificial inflation of scores on teacher-based ODD rating scales.

## Limitations and Strengths

Several limitations to the present study are worth noting. Considerable controversy exists regarding the external validity of findings from analogue studies such as this. The teachers in this study viewed only a small segment of the child's behavior. In addition, during the study the child had the teacher's undivided attention, which rarely occurs during the course of the school day. Though several measures were taken to try to make the classroom appear authentic, it would be naive to believe that this objective was completely met. Several of the teachers commented that, with the exception of the target-child, the children in the classrooms were the most well-behaved and quiet they had ever seen. Unfortunately, the study did not systematically solicit teacher comments regarding the believability of the tapes.

The greatest strength of this study was the degree of internal validity obtained by using an analogue design. Though the teachers may have detected that something was amiss with the classrooms, their responses on the questionnaires still suggested the presence of a halo effect. The degree of control exercised in the study allowed for the inference of cause and effect relationships.

## Conclusions

When disruptive behaviors occurred, teachers in the present study failed to differentiate between the oppositional defiant behaviors associated with ODD, and the hyperactive, inattentive, and impulsive behaviors associated with ADHD. In addition, the magnitude of this failure was a function of the child's gender and the type of behaviors exhibited. Since prevalence studies have frequently relied exclusively on teacher-based rating scales (see, Pelham,

Nagy, Grenslade, & Milich, 1992; Szatmari et al., 1989; Trites, Dugas, Lynch, & Ferguson,1979; Wolraich, Hannah, Pinnock, Baumgaertel, & Brown, 1996), findings from the present study suggest three major implications: (a) compared to females, the male prevalence rate for ADHD may be artificially inflated by the presence of oppositional and defiant behaviors; (b) compared to males, the female prevalence rate for ODD may be artificially inflated by the presence of hyperactive, inattentive, and impulsive behaviors; and (c) the comorbidity rate between ADHD and ODD may be artificially inflated by the failure of teachers to differentiate between the different types of disruptive behaviors. In addition, since similar amounts of bias occurred with both the CTRS-R:S and the DBD, results from the present study suggest that these biases found in teacher-based ratings may not be as attributable to qualities of the rating scales as previously had been hypothesized (Abikoff et al., 1993; Hinshaw, 1987; Schachar et al., 1986, Stevens et al., 1998).

Furthermore, teacher education and greater exposure to materials about disruptive behavior disorders do not seem to be the answer to attenuating these biases. In fact, Stevens et al. (1998) found just the opposite: teachers who had been exposed to more information about ADHD tended to rate the child with ODD as being more inattentive and hyperactive compared to teachers with less exposure to ADHD. Hancock (1996) suggested that the heavy media coverage recently garnered by ADHD has actually contributed to an over- diagnosis of the disorder.

Teachers are a valuable source of information. They are not, however, diagnosticians, nor are they mental health professionals. One should not be surprised, therefore at the biases found in teacher-based ratings of disruptive

behavior disorders, nor should one exclude teacher ratings from the assessment process. Instead, a multi-method assessment approach is recommended for the diagnosis of ADHD (Barkley, 1998; Schaughency & Rothlind, 1991). Ideally, this assessment should include a medical examination, clinical interview of the child, diagnostic interview with the parents, the completion of behavior ratings scales by the parents and teachers, direct observation of the child's behavior by a trained observer, and the administration of clinic-based tests. Barkley (1997) suggested that an adequate ADHD assessment could take up to a year. Based on the above recommendation, the teacher completed-rating scale represents only a small portion of the information to be collected. It behooves mental health professionals to collect and synthesize all necessary information before rendering diagnoses; yet, time and again, both prevalence studies and clinicians have relied almost solely on teacher-based rating scales. Ultimately, a valid diagnosis may depend more on what the clinician does with the rating scale after it has been completed by the teacher than on what the teacher had done with the rating scale.

One additional area for further research that has not yet been discussed involves delineating the factors or mechanisms underlying the biases. It is possible that while engaging in oppositional defiant behaviors, the child with ODD is being inattentive, while appearing hyperactive and impulsive. Likewise, the inattentive, hyperactive and impulsive child with ADHD causes classroom disruptions, which the teacher may view as being oppositional and defiant. Abikoff et al. (1993) suggested that a bidirectional rating bias could result from the "influence of teacher's implicit personality theories regarding disruptive children" (p. 529). Due to the high comorbidity among the disruptive behavior

disorders, teachers may come to assume that hyperactivity and defiance co-occur. As a result, teachers' ratings may reflect these assumptions and expectations rather than the actual behavior of the child. Abikoff et al., however, ruled out this explanation due to the unidirectional nature of the bias found in their study. Since the present study found evidence for a bidirectional bias between ODD and ADHD-type behaviors, Abikoff's implicit personality theory explanation seems possible.

A final possible factor may involve the influence of specific types of disruptive behaviors. The ODD tapes in the present study primarily contained verbal aggression towards the teacher. It may be that this particular behavior is extremely salient for teachers, thus capturing their attention and increasing their vigilance towards the detection of further behavioral problems, possibly at the expense of objectivity. Likewise, 35% of the Hyperactivity Composite scores for the ADHD tapes in the present study involved the "interference" category. Such interfering behaviors may be construed by the teacher as being oppositional and defiant in nature. Future research could, therefore, evaluate the impact of other specific types of oppositional and defiant behaviors on teachers' ratings of ADHD, as well as the impact of other specific ADHD-type behaviors on teachers' ratings of ODD.

# APPENDICES

# APPENDIX A

**Consent to Participate in Research**

My name is David Jackson. I am a graduate student working on my dissertation under the supervision of Dr. Alan King, at the University of North Dakota, Psychology Department. We are conducting a study examining elementary school teacher's perceptions of child behavior.

Basis for participant selection: Any teachers of grades K-6 (general and/or special education) are invited to participate.

Description of study: You will be shown two, ten-minute video taped segments of a fourth grade class involved in a lesson. At the start of each tape, your attention will be drawn to a particular child in the class. After viewing each tape you will be given a 55-item, behavioral questionnaire to complete on the highlighted child. Total time for the study should be approximately one hour. Upon completion of the questionnaires, there will be a debriefing to assure an understanding of the research being conducted.

Participant's Rights: Participation is voluntary. You may choose to not participate, or to discontinue participation at any time without penalty. The decision to not participate will in no way impact your relationship with the University of North Dakota. All responses will be kept strict confidence and there will be no way to trace you to your response sheet. Information obtained will be used solely for research purposes.

Benefits: Each participant will receive a check for $25.00, regardless of whether or not they adequately complete the questionnaire. In addition, participants may request a copy of the results from the study by calling Dr. Alan King at (701) 777-3644.

Potential Risks: Every effort will be taken to minimize the potential for harm or injury in this study. However, in the event that this research activity results in any injury or distress, treatment will be available as it is to the general public in similar circumstances. You or your third party payer must provide payment for any such services, and the University of North Dakota will not be liable.

The investigators involved in this research project will be available to answer any questions you have concerning this program. You may contact David A. Jackson at (701) 777-4348 or Dr. Alan King at (701) 777-3644. You will be given a copy of this form for your own records.

**********

I have read the above consent form and understand my rights as a participant. By signing below, I indicate that I freely choose to participate in this study.

_____ _____

Participant's Signature Date

_____

Print Participant's Name

# APPENDIX B

## CONSENT TO PARTICIPATE IN RESEARCH

My name is David Jackson. I am a graduate student working on my dissertation under the supervision of Dr. Alan King, at the University of North Dakota, Department of Psychology. We are conducting a study examining the impact of student gender on teacher-completed rating forms assessing Attention-Deficit Hyperactivity Disorder, and Oppositional Defiant Disorder.

Basis for participant selection: Children selected for participation in this study must be between the ages of 8 and 10.

Description of study: Your child is being asked to participate in the development of six, ten minute videotapes, which depict a lesson being taught to a fourth grade classroom. The children and teacher will be actors and your child may or may not have speaking lines. The segments have been scripted so that one child in each tape will portray behaviors typical of a child with either Attention-Deficit Hyperactivity Disorder, or Oppositional Defiant Disorder. Participation will require the child's presence at one Saturday and one Sunday taping session, with the possibility of one additional Saturday session. It is anticipated that the first Saturday session will last approximately five hours, while the Sunday session will last approximately four hours. The additional Saturday session, if necessary, will last approximately 3 hours. In addition, children with speaking lines will be expected to memorize the lines on their own time. The two "target" children, with major speaking lines, will be expected to attend one, two-hour rehearsal. It will be necessary that each child be present for the duration of each taping session.

Use of the tapes: The tapes will be shown to elementary school teachers in the region who will be told they are watching an actual fourth grade classroom. After completing rating scales summarizing the behaviors of the highlighted child, the teachers will be informed that they were actually watching role-plays. The tapes will be used solely for educational purposes (research and training), and the names of the child actors will be kept confidential.

Participant's rights: Participation is completely voluntary. You may choose not to participate or to discontinue participation at any time without penalty. If you choose to participate then later withdraw, you will be paid according to the proportion of time your child has invested in the study.

Participation benefits: Parents of each child portraying the ADHD or ODD "target" children will receive $100.00; parents of each child portraying the normal "target" children will receive $75.00, while parents of children with limited or no speaking lines will receive $50.00. Payment will be rendered upon completion of the tapes. If a child is unable to attend for the duration of the taping, they will be paid according to their total number of hours of filming participation. Lunch and other snacks will be provided on taping days. In addition, the children will gain intangible benefits resulting from teamwork in the creation of such a project.

Potential risks: The children will be acting out behaviors that are generally not acceptable in a classroom setting. Prior to the study, and frequently throughout the program, the students will be reminded of this fact. In addition, participation will require effort to memorize lines, and dedication to several days work.

## APPENDIX B, cont.

Your child may become fatigued, bored, and/or frustrated. During each day of filming, we will take a break every hour, and incorporate fun activities.

Every effort will be taken to minimize the potential for harm or risk in this study; however, in the event that research activities result in any injury or distress, treatment will be available, as it is to the general public in similar circumstances. You or your third party payer must provide payment for any such services, and the University of North Dakota will not be liable

The investigators involved with this research project will be available to answer any questions you may have concerning this program. You may contact David Jackson at (701) 777-9826, or Dr. Allan King at (701) 777-3644. You will be given a copy of this form to keep.

I have read the above consent form and understand my rights, and the rights of my child as a participant. Information and understand my rights as a participant. By signing below, I indicate that I freely choose to participate in this study.

_____ _____

Child's Name Child's Age

_____ _____ _____

Parent Signature Phone Number Date

Which of the following three roles would you and your child prefer? If more than one role would be acceptable, please rank-order them. Please note that if too many children sign up for a role, it will be up to the research team to determine which children get which roles; however, under no circumstances will a child receive a role deemed unacceptable to the parents.

_____ Main role portraying both an Attention-Deficit Hyperactivity Disorder Child and an Oppositional Defiant child. Approximately 5 pages of lines to memorize and pay is $100.00

_____ Role of a child portraying a "normal" student. Approximately two pages of lines to memorize and pay is $75.

_____ Supporting role as a child in the classroom. Either no, or minimal lines to memorize and pay is $50.00

_____ Check here and write your address below if you would like a copy of the results of this study.

Please list any allergies, as your child will be served snacks and a lunch on filming days.

# APPENDIX C

## ADHD Script
## (Ben)

Class: There is a general hum of activity for 10 seconds before the teacher settles them down and begins class.

Teacher: "O.K class, I want you all to turn to page 31 in your workbooks. Today's assignment will be numbers 1-33". Teacher writes this on the board. "I will be walking around helping you on this, please put yesterday's homework on your desk, I'll be looking at this also". Teacher writes something else on the board then begins to circulate room-helping children.

Ben: Fidgets with his desk for 5 seconds.
Works for 15 seconds then says:
"Ms Drake, I did this one yesterday."

Teacher: "No, you did one that looks just like this one, now get back to work."

Ben: All children work for 30 seconds. Ben constantly fidgets in his seat.
He fiddles with his pencil, eventually flinging it at BreAnna, who returns the pencil.
Still fiddling with the pencil, he drops it on the floor and has to get out of his seat to get it.

After picking it up, he wanders over to the pencil sharpener, looking over others' desks as he walks.
He sharpens his pencil for 5 seconds, looks at it then continues to sharpen it until interrupted by the teacher.

Teacher: Glances up after 15 seconds and says, "Ben, you need to return to your seat and finish your work".

Ben: "Okay", then slowly returns to his seat.
All children work quietly for 60 seconds. Ben is constantly fidgeting in his seat.
He loudly erases something.
Looks around the room several times.

APPENDIX C. cont.

He takes some Star Wars cards out of his desk, then turns to Jessica and pulls on her shirt twice to get her attention.

He then says in a loud whisper, "I can trade you some Star Wars cards...I got some new ones." He begins to show the cards to Jessica.

BreAnna B: "Shhhhhh.... I'm trying to work."

Ben: Works for 30 seconds

Scene 2

Ben: Looks around the room for 15 seconds
Taps his pencil on the desk for 15 seconds.
While looking at the teacher says,
"Am I going to work on the computer today?"

Teacher: "Ben, be quiet and get back to work."

Ben: "But am I going to be able to work on the computer today?"

Teacher: "Right now you need to get back to work on your paper and we'll think about computers later."

Ben: "Okay."

All children, including Ben, work for 30 seconds. Ben is bouncing his foot.
Ben erases something then looks at someone talking on the other side of the room.
He erases something again and blows the eraser filings onto the floor.
He gets the attention of BreAnna and says: "You got an extra eraser, mine is all worn out."

BreAnna B: Shakes her head no and says, "No, I've only got one"

Ben: Sits quietly back in his chair.
He gets out of his seat and walks over to the back of the class towards the pencil sharpener.
He pulls out a matchbox car from his pocket, and says to Aaron:
" Look what I've got. I got a bunch of them at home. This is just one of the new ones I got last weekend."

He holds it out for the Aaron to see, then puts it Aaron's desk and begins to loudly push it across his desk.

APPENDIX C, cont.

Aaron: "You're going to get me into trouble."

Teacher: "Ben, put that in your pocket, or I'll have to take it away. I've asked you already to get back in your seat. You need to sit down and stay in your seat and get your work finished."

Ben: "Okay Ms. Drake." Ben then wanders back to his seat.
He stops to swing between two desks.
He stops and stares out the window for 5 seconds.

Teacher: In a raised voice; "Ben."

Ben: "I'm going." He returns to his seat.

Scene 3

Ben: After working for 60 seconds, he picks up a pencil box on his desk and opens and shuts it several times (slightly fidgety throughout)
Work for 30 seconds, and then pick up a pencil and loudly taps it on the desk about 10 times.

He rests his head on his hands and halfway begins writing for 15 seconds
He looks up and watches other children for a 15 seconds, then raises his hand.
"Ms Drake, can I go to the bathroom?"

Teacher: "Ben, you just went ten minutes ago, wait until the end of the period."

Ben: Sights loudly, but not angrily, then returns to work for 30 seconds (slightly fidgety throughout).

Teacher: Walks to Ben's desk and has an appropriate interaction (10 seconds) regarding the assignment then walks away. Ben proudly smiles, then continues to work quietly for 10 seconds.

# APPENDIX D

## ODD Script
## (Ben)

Class: There is a general hum of activity for 10 seconds before the teacher settles them down and begins class.

Teacher: "O.K class, I want you all to turn to page 31 in your workbooks. Today's assignment will be numbers 1-33". Teacher writes this on the board. "I will be walking around helping you on this, please put yesterday's homework on your desk, I'll be looking at this also". Teacher writes something else on the board then begins to circulate room-helping children.

Ben: Slightly fidgety, he works appropriately for 1 minute.
Raising his hand he says: "Ms. Drake, I did this one yesterday."

Teacher: "No, you did one that looks just like this one. Now get back to work."

Ben: Lightly slapping his desk in irritation: "How many problems?"

Teacher: "The whole first page".

Ben: Disappointed: "Oh jeez." Then he abruptly turns to his desk and begins the work.
Calmly works at his seat for 30 seconds.
Quietly erases for a 5 seconds.
Works for 20 seconds then slides down in chair, such that his head rests on the seat.

Teacher: "Ben, sit up."

Ben: Slowly gets up (10 seconds). As he does, he stops to tie his shoes.

Teacher: Angry: "Ben, sit up and finish your work, now!"

Ben: Sits up and begins to work.

Teacher: "Okay, class, you've been doing a great job. Evie, would you please collect the papers and put them on my desk". She then writes something on the board.

APPENDIX D. cont.

Scene 2

Evie: "Yes ma'am". She walks around the room collecting papers. Walks to Johnny's desk and waits for his paper. She then says: "Can I have you paper please".

Ben: "Look, I'm not done, so why don't you just get lost".

Evie: Raises her hand and says: "Teacher, Ben wont give me his paper".

Teacher: "Ben, you need to give your paper to Evie. You need to stop writing listen to the directions for the next assignment".

Ben: In an argumentative tone: "Look, I've only got one more to finish".

Teacher: "Ben, you need to give her your paper now".

Ben: Angry: "Fine, okay, if that's what you want". He shoves his paper at Evie.

Evie: Collects the rest of the papers and takes them to the teacher's desk.

Teacher: "Okay, the instructions are on the board for the next page. Now, everyone pick up their pencils and begin working on the next part."

Scene 3

Ben: Works quietly for 60 seconds (slightly fidgety for the first 10 seconds).
He sneakily looks at the teacher, then returns to his work for 30 seconds.
Once again, he sneakily looks at the teacher then returns to his work for 30 seconds (she is still too close for him to do what he wants).

He takes a car out of his box and look at it for 30 seconds.
He next looks at his box, then looks at his car, then to his box, then to his car (do this four times, you just got a great idea).

He slowly makes a bridge between he and BreAnna's desk with the box. He then loudly runs the car across the bridge until the teacher tells him to stop.

Teacher: From the other side of the room angrily says: "Ben, you need to give me that toy, what is this!"

Ben: Looks up startled, acting annoyed, he says, "I'll put it in my desk, I'll put it away."

## APPENDIX D, cont.

Teacher: Walks across the room to Ben's desk, holds out her hand and says: "Ben, you need to give me that toy!"

Ben: Laughs defiantly.

Teacher: "This is not funny....I think you should get up and put it on my desk." (Pointing to a desk).

Ben: Sits at his desk still turned away from the teacher, looking at his car (he is contemplating whether or not he wants to follow her order). He looks at his car for 15 seconds.

Teacher: "Ben"....

Ben: He gets up to put the car on her desk.
He angrily pushes in his chair, and as he walks past his neighbor's desk, he kicks it, then tosses the car onto the teacher's desk.

He then returns to his desk, slumps in his chair, stares at his teacher for a moment, then picks up his pencil, and returns to work.

Jessica: After 10 seconds, Jessica says, "You know your not supposed to have stuff out on your desk, but you'll get it back this afternoon."

Ben: "Mind your own business you dork, leave me alone and get a life".

Works without incident for 2 minutes, (he is slightly fidgety but not overly so).

Abikoff, Howard (1990) with minor revisions by David A. Jackson 11/22/99

# APPENDIX E

Normal Script
(Michael)

Introduction: *The children are in the room just finishing the spelling period and about to change to the math period.*

Teacher: *She is sitting at her desk with Rachel and Abbie standing beside her. They are recording the girls' test scores in the teacher's record book. She says,*
**"OK, Rachel, I've got it."**

Rachel: *She leans over the teacher's desk to see her score in the record book.*

Rachel & Abbie: *Both girls return to their seat.*

Teacher: *As the girls are returning to their chairs, the teacher says,*
**"There..... OK... That about does it."**

*She gathers up her papers and turns to the class:*
**"Now, everybody, it's time for math. Put away your spelling workbooks and get ready."**

Michael: *Pumps his fist in excitement and says:* **"Yes!"**

Class: *There is a flurry of activity; several children open their desks to put their books away. There is a low hum of conversation throughout the room.*

Abbie: *She stands nest to Ms. Drake's desk and says:*
**"My mom wants me to leave early tomorrow because of a doctor's appointment."**

Teacher: *Bends over a little bit to hear Abbie*

Abikoff, Howard 1990, with minor revisions by David A. Jackson 11/99.

APPENDIX E, cont.

**"That's o.k., honey, you just tell her to send a note in so that you can leave."**

Michael: *While everybody is getting ready for math, he leans over to put his spelling book away in his desk and pulls out his pencil box in order to have his pencil and eraser ready. He leans over to say something quietly to Derek He works diligently on his math for the next minute or so.*

BreAnna F: *She moves quickly to sharpen her pencil before the period begins and returns to her seat.*

Mathew: *He goes to the teacher's desk and takes an eraser, then returns to his seat.*

Teacher: *Through all this Ms. Drakei s answering questions.*

Jordan: *When Ms. Drake gets near him, he raises his hand and quietly asks,*
**"I can't find my pencil. Can I have another?"**

Teacher: *She goes to her desk and gets Jordan a pencil.*

**Instructions:**

Teacher: **"All right, let's get started."**
*She waits for the class to look up at her, Pause.*

**"O.K."** *Pause.*

**"Class, I am going to hand out some math sheets. I want you to practice some more... because the problems that you handed in yesterday still had too many mistakes in them... Remember to subtract from the top down."**

Michael: *Stops working on math and swivels around in his chair to listen to*

APPENDIX E, cont.

*the teacher.*

Teacher: *Pause to look around the room.*
**"There were far TOO MANY errors due to carelessness." I'm going to hand out the first sheet of problems. I want you to work on them quietly. I'm going to be walking around the room helping anybody who needs it. If you have any questions, please raise your hand quietly. When you have finished the first page, raise your hand... THAT DOESN'T MEAN GET OUT OF YOUR SEAT, RIGHT?** (The whole class responds by saying **"right Ms. Drake"** in unison) **Raise your hand and I'll give you the next page."**

*She hands the papers to Naomi to hand out to the class.*

Class: *Everybody begins to work on their math papers.*

Michael: *He puts his name on the top of the paper and begins working.*

*There is a 30 second pause while the children begin working and the teacher begins walking around the room to settle her class down and make sure they've started off on the right foot.*

Teacher: *She moves on to a child.*
**"That's good. You do nice work... Put your name on the top of the page."**

*In a louder voice,*
**"Remember, one of the first rules of the room is to put your names on your work, everybody."**

Class: *Returns to work.*

**Class Works and Naomi Sharpens a Pencil**

Class: *They all return to work for two minutes.*

APPENDIX E, cont.

Teacher: *Refocuses two girls who are talking.*
**"Evie and Jordan, get back to work now."**

Michael: *Acting slightly fidgety, he slides down a bit in his chair as he works. He is trying to figure out a difficult problem.*

Teacher: *Leans over a child's desk in back.*
**"Let's see how you're doing. OK. That's a correct answer for this one...but what do you think about the second one? ... Yes, that's better."**

Class: *Works for two minutes.*

Teacher: *She moves down to the next child and silently watches her work for a few minutes.*
**"Um humm...Okay"**
*She moves to the other side of the table and looks at another child's work. She leans over to point something out.....*

Naomi: *Raises her hand:*
**"Ms. Johnson".**

Teacher: **"Yes Naomi."**

Naomi: **"My pencil is broken, can I get a new one?"**

Teacher: **"No, why don't you try to sharpen that one."**

Naomi: *She goes to the pencil sharpener and sharpens her pencil.*

Teacher: *Moves to look at Mathew's work*
**"Oh look, you forgot to put your last name on the paper."**

*Pause:*
**"All right, that's better"**

*She goes off screen so we just hear her voice.*
**"Let's see how you're doing here at the back table. You've been quiet here."**

APPENDIX E, cont.

*While off screen she says:*
**"Stay within the space for the answer....That's better. Um hum...Good."**

**Class Works and Aarom Gets a Star**

Class: *Works for 1 and 1/2 minutes..*

Michael: *Erases an answer slowly. He slides sideways in his chair and, leaning his head on his hand, he stares off into the distance for 15 seconds. He is daydreaming. Then he goes back to work.*

Aaron: *Raises his hand.*
**"Ms. Drake, I'm finished. What's the next page?"**

Teacher: **"Oh your finished, let me see"**

*She walks over to Aaron's desk.*
**"Um...good. and your name is on the top. Your first page will have a star on it because you did such a good job. Okay, I'll get you the second sheet."**

*Teacher goes to the front of the class get a second sheet for Aaron and returns.*
**"Here you go."**

*She puts the star on.*
**"Write your name on the top of it. Your last name goes here. Umm-humm... Do it a little smaller. Okay."**

Aaron: *Goes back to work.*

Teacher: *Has moved to oversee the work of another child.*
**"Okay, good."**

**Class Works and Michael Gets Help.**

Rachel: *Coughs.*

APPENDIX E, cont.

Teacher: *She leans over table 3 and shows Rachel how to do a problem, her back is mostly turned away from Michael*

Michael: *Turns around in his chair to see where she is and raises his hand.*
**"Ms. Drake, can you help me with this one."**

Teacher: **"OK, I'll be there in just a minute..."**
*She turns back to what she was doing with Rachel.*
**"Name...very good. You kept going at it until you did it, didn't you?"**
*She walks over to Michael*

Michael: **"I can't remember how to do this kind."**

Teacher: *Bends over his deskso that her face can't be seen, but Michael's can be seen.*
**"Let me see..."** *Pause.*

**"You remember, what do you do when the first column adds up to a number greater than ten?"** *Pause.*

*She watches silently as he completes the problem.*
**"Um hum... Right!"**

*Then she walks on to watch another child.*

Class: *Continues to work for 1 minute.*

Michael: *Works slower and slower. He taps his pencil on the desk while he works. It drops and he leans over to get it.*

*When he picks it up, he leans over and says something to Abbie. Then he goes back to work.*

# APPENDIX F

## Classroom Observations

The classroom observation system is an interval recording system that generates data regarding children's behavior in the classroom. Each observer watches three children simultaneously and records the occurrence of disruptive behaviors exhibited by the children. Each observer also records on-task behavior using a time-sampling technique. The data gathered through the use of the classroom observation system are used to evaluate the effects of medication on children's behavior in the classroom.

### Categories and Codes

The seven disruptive behavior categories, the operational definition for each category, and the codes used to represent these categories are listed below. Examples of behaviors that meet the criteria for each category are also listed.

Physical Aggression/Intrusion Code: P

The child performs a physical behavior that (1) would typically produce immediate physical injury or pain to another, or (2) intrudes on another by inappropriately restricting freedom of movement, or (3) otherwise elicits clear behavioral indications of annoyance or distress from the recipient. **Note:** Any act of aggression directed toward another person that would typically result in discomfort is recorded as Physical Aggression/Intrusion, regardless of whether harm occurs.

Examples: Pushing a classmate ... pinching a classmate ... kicking a classmate ... hitting the teacher ... grabbing any part of another's body ... shoving a desk toward another child ... pulling a child by his or her shirt ... tapping a child on the head with a pencil.

Notes:

Verbal Abuse Code: A

The child produces a communication, either vocal or nonvocal, designed to elicit a clear behavioral indication of annoyance or distress from the intended recipient. **Note:** Behaviors such as swearing, name calling, teasing, or threatening are recorded as Verbal Abuse.

Examples: saying hell, asshole, shit, etc. ... calling a classmate a dumb bunny ... sticking out tongue ... "giving the finger" ... making a face at another child or at the teacher ... shaking fist at another person ... saying, "I'm going to tell the teacher" ... making a face at the teacher when the teacher is not looking ... making farting noises with armpit.

Notes:

## APPENDIX F, cont.

Destruction of Property/Inappropriate Use of Materials Code: D

The child destroys or damages an object, or defaces its surface, such that the object's value or usefulness is impaired or reduced at least temporarily. The child uses materials for purposes other than those for which the materials were designed or intended.

Examples: Writing on desk ... breaking a pencil ... tearing a workbook ... scribbling in a book ... throwing a book ... making a paper airplane ... crumbling an assignment ... slamming a book on a table ... hitting, kicking, or tapping desk ... opening desk without permission ... putting any materials into mouth ... rocking or rattling a desk ... chewing on clothing or other items such as rings, necklaces, watches, or buttons ... tapping a pencil on any object in a manner that produces noise ... playing games with a pencil (e.g., using a pencil as an airplane, bomb, or person).

Notes: If a child's materials are on the floor at the start of an observation interval, observers should record a "D" for the current interval only. If the materials remain on the floor during subsequent intervals, observers should not record a "D". If a child is using a pencil as a pointer when reading, using a pencil to count on fingers, or raising hand while holding a pencil, observers should not record a "D".

Cheating Code: C

The child (1) obtains information concerning academic tasks related to that child to which access is not permitted, or (2) gives information that the child should not give.

Examples: Looking at another paper while working on an assignment ... copying answers from a book when the teacher is not looking ... beginning to work before the teacher gives permission ... telling another child the answer to a problem ... looking at another child's paper.

Notes:

Verbal Intrusion Code: I

The child produces a vocalization, with or without meaning, that (1) intrudes into the activity or conversations of others, (2) interrupts another person's currently assigned activity or occurs at the same time the other is talking, or (3) is distracting or intrusive without the aggressive or abusive quality of verbal abuse. The vocalization must have an intended recipient. **Note:** A behavior that indicates a communication but cannot be heard is recorded as Verbal Intrusion.

Examples: Calling out without permission ... interrupting another child who is talking ... talking to another child without permission ... responding, "thanks," to a teacher who says "keep up the good work" (a response is not typically expected).

Notes:

## APPENDIX F, cont.

Talking to Self — Code: T

The child produces vocalizations, with or without meaning, in the absence of another person who is identifiable as the intended recipient of the communication. The vocalization must be loud enough to be heard.

Examples: [all in the absence of an identifiable recipient] Laughing without an apparent stimulus ... saying oops, all right!, oh no!, etc. ... talking, humming, singing, or whistling while doing seatwork ... saying digits aloud while working on math problems ... reading aloud ... yawning ... tapping feet or fingers ... crying ... producing noise without the use of an object (e.g., clapping hands) ... saying "Ouch!" in response to prodding by another child.

Notes:

Out-of-Seat — Code: O

The child is out of his or her seat without permission.

Examples: Standing up ... leaving chair without permission.

Notes:

Attending; On-Task Behavior — Code: +/-

The child attends to the current assigned task. Attention must be indicated by: (1) looking at or manipulating objects or materials on the child's desk that are necessary for completing the task, (2) looking at the blackboard or another location where materials related to the task are displayed, (3) looking at an instructor who is in the process of providing instruction regarding the task, (4) looking at any object or place to which the child has been directed by the instructor, (5) looking at a peer who has been asked a question by the instructor during a group lesson, (6) performance of a motor activity as required by the task, or (7) performance of a motor activity for the purpose of preparing for or finishing a current assigned activity.

Examples: Sitting at desk writing an assignment ... looking at the teacher while the teacher is talking ... looking at blackboard for instructions to a task ... looking at a map on the wall to which the teacher is pointing during a lesson ... looking at a peer who is answering a question in group instruction ... passing out papers before a math assignment as requested by teacher ... looking at or counting on fingers while doing a math assignment ... looking at aide who is commenting on child's work ... looking at another child who is responding to the teacher during a lesson ... looking straight ahead after finishing all assignments ... looking at the teacher while raising hand to ask a question.

Notes: Shuffling through papers is off-task.

APPENDIX F, cont.

**Procedures**

Before entering the classroom, observers should gather all the necessary materials and should test all equipment. The materials needed are a clipboard, three classroom observation data sheets, a pencil, a tape recorder with an earphone, an AC adapter, and a prerecorded beeper tape. Observers should also make sure that the tape is rewound and the player is working with sufficient time before the beginning of the ALC period to get a new tape, adapter, or recorder if needed. Observers should enter the classroom at least five minutes before the beginning of the period and should set up and test all necessary equipment before the children enter the classroom. The classroom should be arranged such that children sit in four rows of three children each. Observers should sit in the front corner of the classroom in a manner that allows the observers to see and hear all the children but does not distract the children or the learning center staff members. Observers should not talk to the children or make eye contact with the children during the period.

The developmental specialist may use the first few minutes of the period to review the Academic Learning Center (ALC) rules and procedures with the children or to inquire about special group activities such as field trips or individual rewards such as High Point Kid. Observers should not record data during this part of the ALC period. At the end of this discussion, the developmental specialist tells the children to begin working on their assignments and sets the classroom timer for 45 minutes. Observers should begin recording data when the developmental specialist tells the children to begin working on their assignments. Observers should continuously observe the children and record data for the entire 45-minute work period, and should stop recording data when the developmental specialist tells the children to stop working on their assignments. Observers should not record data while the developmental specialist is providing feedback to the children at the end of the period. However, observers should not talk, stand, make noise, leave the classroom, or otherwise distract the children or the ALC staff members during the feedback session.

The observation interval used in the classroom observation system is a 20-second interval. During the first 15 seconds of the interval, observers look for and record the occurrence of disruptive behaviors. During the last 5 seconds of the interval, observers evaluate and record on-task behavior. Observers listen to a prerecorded cassette tape that announces the beginning of the 15-second period and the beginning of the 5-second period. At the end of one interval, observers shift their attention to another group of children and begin the cycle again.

During the first 15 seconds of the interval, each observer watches three children simultaneously. For each child, the observer records the first occurrence of any behavior that meets the criteria for any of the seven disruptive behaviors described above. Observers may record only one disruptive behavior for each child during an interval. For example, if a child scribbles on his desk and then burps loudly, the observer should record the occurrence of Destruction of Property but should not record the occurrence of Verbal Intrusion. If a child does not exhibit a disruptive behavior during the observation interval, the observer should record a dash (—) in the behavior portion of that child's data box for

that interval.

If a child is absent or is out of the classroom for any reason, the observer should not change the observation groupings. That is, the observer should observe the remaining children in that child's row, and should not add a third child from the next row.

If a child is serving a time out during an interval and can therefore not be observed, the observer should record "TO" in the behavior portion of the recording box at the end of the interval. If a child is absent for any other reason (e.g., late arrival, illness), the observer should record an asterisk in the behavior portion of the recording box. It is possible for a child to have behavioral data for an interval, but to be out of the classroom when the observer records on-task behavior (e.g., a child could display aggressive behavior at the beginning of the interval and could be serving a time out at the time the observer observes on-task behavior. In this situation, the observer should record an asterisk in the on-task portion of the recording box.

At the beginning of the five-second on-task recording period, observers should glance briefly at each of the three children being observed and should record whether each child's behavior meets the criteria for on-task behavior at that moment. After recording on-task behavior, the observers should prepare to observe the next three children listed on the data sheet.

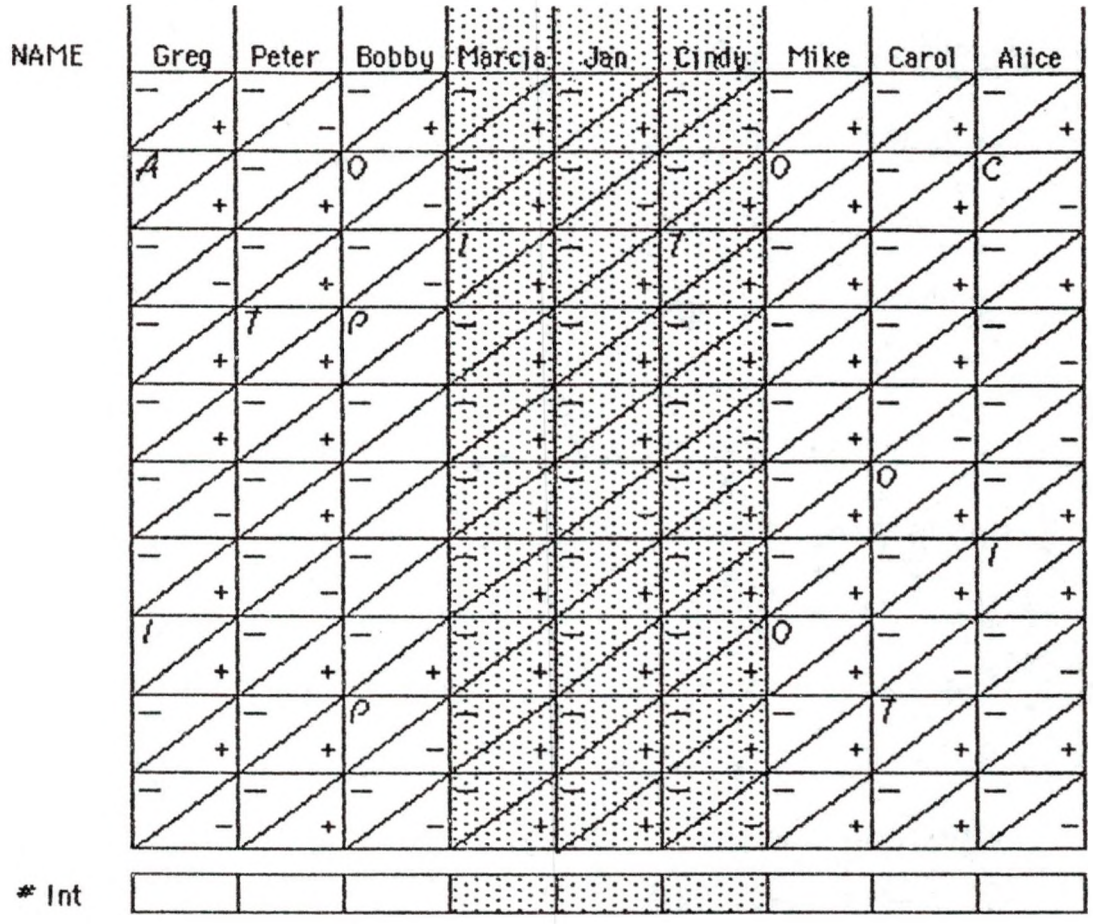
**Classroom Observation Data Sheet**

Below is a Classroom Observation Data Sheet with sample data. The shaded columns indicate triads of children that observers should observe simultaneously. Each row of the data sheet indicates a new series of observations for each triad. To facilitate data recording, each box of the data sheet is split by a diagonal line. Observers should record disruptive behavior codes in the upper-left corner of the data boxes and should record on-task behavior codes in the lower-right corners of the data boxes.

At the beginning of the observation period, observers should observe the three children listed in the first three columns of the data sheet and should record data in the corresponding boxes of the first row. Observers should then observe the next three children listed on the data sheet and should record data for these children in the corresponding boxes. After recording data in the last three columns of the first row, observers should observe the first triad of children again and should record data in the corresponding boxes of the second row. Observers should continue to observe triads of children and to record data until the end of the seatwork period.

APPENDIX F, cont.

DATE: 7/14 DAY #: 26 GROUP: 4 OBSERVER: TP




**Scoring**

Observers calculate the percentage of intervals during which each child exhibited disruptive behaviors and the percentage of intervals during which each child exhibited on-task behavior for use in the data management system. In addition to calculating these percentages for the entire seatwork period, observers calculate these percentages separately for the first half and second half of the seatwork period. To summarize the data and to facilitate the calculation of these percentages, observers use the Classroom Observation Summary Sheet. A Classroom Observation Summary Sheet with sample data follows.

To prepare classroom observation data for computer entry, observers must first determine the midpoint of the seatwork period. This midpoint is used to divide the observation period and is the same for all children regardless of the number of intervals an individual child was observed. To determine the midpoint, observers should count the number of rows that contain data, should divide this number by two, and should draw a line on the data sheet to separate the observation period into halves. Observers should then summarize each child's data and should calculate percentages for the first half, second half, and total observation period.

APPENDIX F, cont.

DATE: 7/14 DAY #: 26 GROUP: 4 OBSERVER: TP

| First Half | Greg | Peter | Bobby | Marcia | Jan | Cindy |
|---|---|---|---|---|---|---|
| Physical Intrusion/Aggression | 0 | 0 | 1 | 0 | 0 | 0 |
| Verbal Abuse | 1 | 0 | 0 | 0 | 0 | 0 |
| Destruction/Inappropriate Use | 0 | 0 | 0 | 0 | 0 | 0 |
| Cheating | 0 | 0 | 0 | 0 | 0 | 0 |
| Verbal Intrusion | 0 | 0 | 0 | 1 | 0 | 0 |
| Talking to Self | 0 | 1 | 0 | 0 | 0 | 1 |
| Out of Seat | 0 | 0 | 1 | 0 | 0 | 0 |

| | | | | | | |
|---|---|---|---|---|---|---|
| # Disruptive First Half | 1 | 1 | 2 | 1 | 0 | 1 |
| # Intervals First Half | 5 | 5 | 4 | 5 | 5 | 5 |
| Percentage Disruptive First Half | 20 | 20 | 50 | 20 | 0 | 20 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Intervals On-task First Half | 4 | 4 | 1 | 5 | 4 | 3 |
| # Intervals First Half | 5 | 5 | 3 | 5 | 5 | 5 |
| Percentage On-task First Half | 80 | 80 | 33 | 100 | 80 | 60 |

APPENDIX F, cont.

| Second Half | Greg | Peter | Bobby | Marcia | Jan | Cindy |
|---|---|---|---|---|---|---|
| Physical Intrusion/Aggression | 0 | 0 | 0 | 0 | 0 | 0 |
| Verbal Abuse | 0 | 0 | 1 | 0 | 0 | 0 |
| Destruction/Inappropriate Use | 0 | 0 | 0 | 0 | 0 | 0 |
| Cheating | 0 | 0 | 0 | 0 | 0 | 0 |
| Verbal Intrusion | 1 | 0 | 0 | 0 | 0 | 1 |
| Talking to Self | 0 | 0 | 0 | 0 | 1 | 0 |
| Out of Seat | 0 | 0 | 0 | 1 | 1 | 0 |
| # Disruptive Second Half | 1 | 0 | 1 | 1 | 2 | 1 |
| # Intervals Second Half | 5 | 5 | 3 | 5 | 5 | 5 |
| Percentage Disruptive Second Half | 20 | 0 | 33 | 20 | 40 | 20 |
| Intervals On-task Second Half | 3 | 4 | 1 | 4 | 4 | 4 |
| # Intervals Second Half | 5 | 5 | 3 | 5 | 5 | 5 |
| Percentage On-task Second Half | 60 | 80 | 33 | 80 | 80 | 80 |

| Total | Greg | Peter | Bobby | Marcia | Jan | Cindy |
|---|---|---|---|---|---|---|
| Total # Disruptive | 2 | 1 | 3 | 2 | 2 | 2 |
| Total # Intervals | 10 | 10 | 7 | 10 | 10 | 10 |
| Total Percentage Disruptive | 20 | 10 | 43 | 20 | 20 | 20 |
| Total Intervals On-task | 7 | 8 | 2 | 9 | 8 | 7 |
| Total # Intervals | 10 | 10 | 6 | 10 | 10 | 10 |
| Total Percentage On-task | 70 | 80 | 33 | 90 | 80 | 70 |

After determining the midpoint of the observation period, observers calculate Percentage Disruptive for the first half of the observation period. Observers should count the number of intervals that each child was observed during the first half of the observation period and should record these numbers in the appropriate boxes of the summary sheet. Observers should then count the occurrences of each disruptive behavior category during the first half and should record these numbers in the appropriate boxes of the summary sheet. Observers should calculate the total number of disruptive behaviors that occurred by adding the frequencies of the individual disruptive behavior categories. To calculate Percentage Disruptive for the first half of the observation period, observers should divide

the total number of disruptive behaviors by the number of intervals that the child was observed and should record this number in the appropriate box of the summary sheet.

To determine Percentage On-task for the first half of the observation period, observers should first count and record the number of intervals for which On-task was evaluated. Observers should then count and record the number of intervals for which a plus sign was recorded. To calculate Percentage On-task, observers should divide the total number of intervals during which the child was on-task by the number of intervals that the child was observed and should record this number in the appropriate box of the summary sheet.

Observers should repeat the procedures described above to calculate Percentage Disruptive and Percentage On-task for the second half of the observation period. Observers should record these numbers in the corresponding boxes of the summary sheet.

After calculating Percentage Disruptive and Percentage On-task for the first and second halves of the observation period, observers should calculate these percentages for the observation period as a whole. Observers should add the number of disruptive behaviors from the first half to the number of disruptive behaviors from the second half and should record the total number of disruptive behaviors for each child in the appropriate boxes of the summary sheet. Observers should total the number of intervals for which disruptive behaviors were evaluated from the two halves of the observation period and should record these numbers in the appropriate boxes of the summary sheet. To calculate Total Percentage Disruptive, observers should divide the total number of disruptive behaviors by the total number of intervals and should record these numbers in the appropriate boxes of the summary sheet. Observers should repeat this procedure to calculate Total Percentage On-Task.

If a child is not present for a substantial portion of one of the halves of the observation period, the child's data cannot be interpreted for that half of the period. Therefore, when scoring the data for each half, if a child was not present for at least half of those intervals, the observer should record asterisks in the Percentage Disruptive and the Percentage On-task rows for that child. The observer should use all available data when computing the total percentages for the child, however.

**Reliability**

Two observers simultaneously observe approximately 20 percent of the Academic Learning Center periods and independently record behavioral data. The data from these observations are used to calculate reliability coefficients. Observers meet regularly to discuss the reliability data and to resolve any problems that occur.

Observers should not talk to each other or look at each other's data sheets while conducting reliability observations. However, it may be necessary for observers to confirm periodically that they are observing the same children. To avoid disrupting the children or the learning center staff members, observers should whisper or use hand signals to synchronize observations.

Entering Reliability Data

Observers enter reliability data into a Microsoft Excel spreadsheet that calculates Cohen's Kappa, a reliability statistic that takes into account the number of intervals during which observers agreed, the number of intervals during which observers disagreed, and frequencies of chance.

## APPENDIX F, cont.

A portion of the entry file is depicted below. For each row of data, observers should enter across the sheet such that there are 12 children's data for each row, as illustrated. If the observer did not record a disruptive behavior or recorded off-task, he or she should enter a zero. If the observer recorded any of the disruptive behaviors (it is not necessary to distinguish between the behaviors) or recorded on-task, he or she should enter a one. If there is missing data for a child or for a row, the observer should enter an asterisk in the cell. In addition, if there is extra space at the bottom of the file after all data have been entered, the observer should enter asterisks for the remaining rows. After entering the observer data, the observer should then enter the reliability data. A second research assistant should verify all data entry. The spreadsheet will automatically calulate Kappa values. The observer should print the summary section of the file and should file the report in the Reliability folder.

# APPENDIX G

## COCADD SCORING PROTOCOL

Name:_____ Date:_____ Tape#:_____

| Category A | | Category B | |
|---|---|---|---|
| Physical Aggression / Intrusion | P | Verbal Intrusion | I |
| Verbal Abuse | A | Inappropriate Use of Materials | M |
| Destruction of Property | D | Talking to Self | T |
| Cheating | C | Out-Of-Seat | O |

Attending: On-Task Behavior Y/N

| Period 1 | Period 2 | Period 3 |
|---|---|---|
| Category A: P A D C | Category A: P A D C | Category A: P A D C |
| Category B: I M T O | Category B: I M T O | Category B: I M T O |
| Attending: Y N | Attending: Y N | Attending: Y N |

| Period 4 | Period 5 | Period 6 |
|---|---|---|
| Category A: P A D C | Category A: P A D C | Category A: P A D C |
| Category B: I M T O | Category B: I M T O | Category B: I M T O |
| Attending: Y N | Attending: Y N | Attending: Y N |

| Period 7 | Period 8 | Period 9 |
|---|---|---|
| Category A: P A D C | Category A: P A D C | Category A: P A D C |
| Category B: I M T O | Category B: I M T O | Category B: I M T O |
| Attending: Y N | Attending: Y N | Attending: Y N |

| Period 10 | Period 11 | Period 12 |
|---|---|---|
| Category A: P A D C | Category A: P A D C | Category A: P A D C |
| Category B: I M T O | Category B: I M T O | Category B: I M T O |
| Attending: Y N | Attending: Y N | Attending: Y N |

| Period 13 | Period 14 | Period 15 |
|---|---|---|
| Category A: P A D C | Category A: P A D C | Category A: P A D C |
| Category B: I M T O | Category B: I M T O | Category B: I M T O |
| Attending: Y N | Attending: Y N | Attending: Y N |

APPENDIX G, cont.

| Period 16 | Period 17 | Period 18 |
|---|---|---|
| Category A: P A D C | Category A: P A D C | Category A: P A D C |
| Category B: I M T O | Category B: I M T O | Category B: I M T O |
| Attending: Y N | Attending: Y N | Attending: Y N |
| **Period 19** | **Period 20** | **Period 21** |
| Category A: P A D C | Category A: P A D C | Category A: P A D C |
| Category B: I M T O | Category B: I M T O | Category B: I M T O |
| Attending: Y N | Attending: Y N | Attending: Y N |
| **Period 22** | **Period 23** | **Period 24** |
| Category A: P A D C | Category A: P A D C | Category A: P A D C |
| Category B: I M T O | Category B: I M T O | Category B: I M T O |
| Attending: Y N | Attending: Y N | Attending: Y N |
| **Period 25** | **Period 26** | **Period 27** |
| Category A: P A D C | Category A: P A D C | Category A: P A D C |
| Category B: I M T O | Category B: I M T O | Category B: I M T O |
| Attending: Y N | Attending: Y N | Attending: Y N |
| **Period 28** | **Period 29** | **Period 30** |
| Category A: P A D C | Category A: P A D C | Category A: P A D C |
| Category B: I M T O | Category B: I M T O | Category B: I M T O |
| Attending: Y N | Attending: Y N | Attending: Y N |

## REFERENCES

Abikoff, H., Courtney, M., Pelham, W. E., & Koplewicz, H. S. (1993). Teachers' ratings of disruptive behaviors: the influence of halo effects. Journal of Abnormal Child Psychology, 21 (5), 519 - 533.

Abikoff, H., Gittleman-Klein, R., & Klein D. F. (1977). Validation of a classroom observation code for hyperactive children. Journal of Consulting and Clinical Psychology, 45 (5), 772 - 783.

Abramwitz, C. V., & Dokecki, P. R. (1977). The politics of clinical judgment: Early empirical returns. Psychological Bulletin, 84, 460-476.

Achenbach, T. M. (1978). The child behavior profile: I. Boys aged 6-11. Journal of Consulting and Clinical Psychology, 46, 478-488.

Achenbach, T. M. (1991). Manual for the Child Behavior Checklist and Revised Child Behavior Profile. Burlington, VT: University Associates in Psychiatry.

Achenbach, T. M., Bird, H. R., Canino, G., et al. (1990). Epidemiologic al comparisons of Puerto Rican and U.S. mainland Children: parent, teacher, and self-reports. Journal of the American Academy of Child and Adolescent Psychiatry, 29, 84 - 93.

American Psychiatric Association (1957). Diagnostic and statistical manual of mental disorders. Washington, DC: American Psychiatric Association.

American Psychiatric Association (1968). Diagnostic and statistical manual of mental disorders (2nd ed.). Washington, DC: American Psychiatric Association.

American Psychiatric Association (1980). Diagnostic and statistical manual of mental disorders (3rd ed.). Washington, DC: American Psychiatric Association.

American Psychiatric Association (1987). Diagnostic and statistical manual of mental disorders, (3rd ed., rev.). Washington, DC: American Psychiatric Association.

American Psychiatric Association (1994). Diagnostic and statistical manual of mental disorders (4th ed.). Washington, DC: American Psychiatric Association.

Anderson, J. C., Williams, S., McGee, R., & Silva, P. A. (1987). DSM-III disorders in preadolescent children. Archives of General Psychiatry, 44, 69-76.

Arcia, E., & Conners, K. (1998). Gender differences in ADHD? Journal of Developmental and Behavioral Pediatrics, 19 (2), 77-83.

Arnold, L, E. (1996). Sex differences in ADHD: Conference summary. Journal of Abnormal Child Psychology, 24 (5), 555-569.

Atkins, M. S., Pelham, W. E., & Licht, M. H. (1985). A comparison of objective classroom measures and teacher ratings of attention deficit disorder. Journal of Abnormal Child Psychology, 13 (1), 155 - 167.

Atkins, M. S., Pelham, W. E., & Licht, M. H. (1988). The development and validation of objective classroom measures for the assessment of conduct and attention deficit disorders. In R. Prinz (Ed.), Advances in behavioral assessment of children and families (Vol. 4, pp. 3-33). Greenwich CT: JAI Press.

Atkins, M. S., Pelham, W. E., & Licht, M. H. (1989). The differential validity of teacher ratings of inattention / overactivity and aggression. Journal of Abnormal Child Psychology, 17 (4), 423 - 435.

August, G. J., Realmuto, G.M., MacDonald, A. W , Mugent, S. M., & Crosby, R. (1996). Prevalence of ADHD and comorbid disorders among elementary school children screened for disruptive behavior. Journal of Abnormal Child Psychology, 24 (5), 571 - 59.

Ball, R. S., Merrifield, P., & Scott, L. H. (1978). Extended Merril-Palmer Scale. Chicago: Stoelting.

Barkley, R. A. (1989). Hyperactive girls and boys: Stimulant drug effects on mother-child interactions. Journal of Child Psychology and Psychiatry, 30 (3), 379-390.

Barkley, R. A. (1990). Attention-deficit hyperactivity disorder: A handbook for diagnosis and treatment. New York: Guilford.

Barkley, R. A. (1994). Attention deficit hyperactivity disorder: A handbook for diagnosis and treatment (3rd ed.). New York: Guilford Press.

Barkley, R. A. (1995). Sex differences in ADHD. ADHD Report, 3 (1), 1-4.

Barkley, R. A. (1997 ). Understanding ADHD and self control. In R. A. Barkley (Ed.). ADHD and the nature of self control. New York: Guilford.

Barkley, R. A., & Edelbrock, C. S. (1987). Assessing situational variation in children's problem behaviors: The Home and School Situations Questionnaires. In R. J. Prinz (Ed.), Advances in behavioral assessment of children and families (Vol. 3, pp. 157-176). Greenwich, CT : JAI.

Bauermeister, J. J. (1992). Factor analyses of teacher ratings of attention-deficit hyperactivity and oppositional defiant symptoms in children

aged four through thirteen years. Journal of Clinical Child Psychology, 21, 27-34.

Berry, C. A., Shaywitz, S. E., & Shaywitz, B. A. (1985). Girls with attention deficit disorder: A silent minority? A report on behavioral and cognitive characteristics. Pediatrics, 76 (5), 801 - 809.

Biederman, J., Faraone, S. V., Spencer, T., Wilens, T. & Mick, E. (1994). Gender differences in adults with attention deficit hyperactivity disorder. Psychopharmacology Bulletin, 30, 653.

Block, G. H. (1977). Hyperactivity: A cultural perspective. Journal of Leaning Disabilities, 110, 236-240.

Bradley, C. B. (1937). The behavior of children receiving benzedrine. American Journal of Psychiatry, 94, 577-585.

Breen, M. J. (1989). Cognitive and behavioral differences in ADHD boys and girls. Journal of Child Psychology and Psychiatry, 30 (5), 711-716.

Brito, G. N. O., Pinto, R. C. A., & Lins, M. F. C. (1995). A behavioral assessment scale for attention deficit disorder in Brazilian children based on DSM-III-R criteria. Journal of Abnormal Child Psychology, 23, 509-520.

Brown, L. S. (1992). A feminist critique of the personality disorders. In L.S. Brown & M. Ballou (Eds.), Personality and psychopathology: Feminist reappraisals (pp. 206-228). New York: Guilford Press.

Brown, R. T., Mandan-Swain, A., Baldwin, K. (1991). Gender differences in a clinic referred sample of attention deficit disorder children. Child Psychiatry and Human Development, 22 (2), 111-128.

Caplan, P. J. (1995). They say you're crazy. How the world's most powerful psychiatrists decide who's normal. Reading, MA: Addison-Wesley.

Carlson, C. L., Pelham, W. E., Milich, R. & Dixon, J. (1992). Single and combined effects of methylphenidate and behavior therapy on the classroom performance of children with attention-deficit hyperactivity disorder. Journal of Abnormal Child Psychology, 20 (2), 213-232.

Carlson, C. L., Tamm, L., & Gaub, M. (1997). Gender differences in children with ADHD, ODD, and co - occurring ADHD / ODD identified in a school population. Journal of the American Academy of Child and Adolescent Psychiatry, 36 (12), 1706 - 1714.

Carter, C. O. (1969). Genetics of common disorders. British Medical Bulletin, 25, 52-57.

Clements, S. D. (1966). Task Force One: Minimal brain dysfunction in children. (National Institute of Neurological Diseases and Blindness, Monograph No. 3). Rockville, MD: U.S. Department of Health, Education, and Welfare.

Clonninger, G. R., Christiansen, K. O., Reich, T., & Gottesman, I. L. (1978). Implications of sex differences in the prevalence of anti-social personality, alcoholism, and criminality for familial transmission. Archives of General Psychiatry, 35, 941-951.

Cohen, J. (1977). Statistical Power analysis for the behavioral sciences. New York: Academic Press.

Cohen, P., Velez, N., Kohn, M., Schwab-Stone, M., Johnson, J. (1987). Child psychiatric diagnosis by computer algorithm. Journal of the Academy of American Child and Adolescent Psychiatry, 26, 631-638.

Conners, C. K. (1969) A teacher rating scale for use in drug studies with children. American Journal of Psychiatry, 126, 884-888.

Conners, C. K. (1973). Rating scales for use in drug studies with children. Psychopharmacology Bulletin, 9, 24-29.

Conners, K. C. (1989). Conners' Rating Scale Manual. Toronto: Multi-Health Systems, Inc.

Conners, K. C. (1997). Conners' Rating Scales Revised: Technical manual. New York: Multi-Health Systems, Inc.

Cooper, P., & Ideus, K. (1995). Is attention deficit disorder a Trojan Horse? Behaviour, 10 (1), 29 - 34.

Cotungo, A. J. (1993). The diagnosis of attention deficit hyperactivity disorder (ADHD) in community mental health centers: Where and when. Psychology in the Schools, 30, 338-344.

DaHass, P. A. (1986). Attention styles and peer relationships of hyperactive and normal boys and girls. Journal of Abnormal Child Psychology, 14 (3), 457-467.

Daley, C. E., Onwuegbuzie, A. J., & Griffin, H. (1998). Attention-deficit/hyperactivity disorder: Relations between prevalence rate and school district size, diagnostic method, and referral process. Psychological Reports, 83 (2), 593-594.

DeFries, J. (1989). Gender ratios in children with reading disability and their affected relatives: A commentary. Journal of Learning Disabilities, 22, 543-545.

Desgranges, K., Desgranges, L., & Karsky, K. (1995). Attention deficit disorder: problems with preconceived diagnosis. Child and Adolescent Social Work Journal, 12 (1), 3-17.

Dunn, L. M., & Dunn, L. M. (1981). Peabody Picture Vocabulary Test-Revised. Circle Pines, MN: American Guidance Service.

DuPaul, G. R. (1991). Parent and teacher ratings of ADHD symptoms: Psychometric properties in a community based sample. Journal of Clinical Child Psychology, 20 (4), 242-253.

Eme, R. F. (1992). Selective female affliction in development of disorders of childhood: A literature review. Journal of Clinical Child Psychology, 21 (4), 354-364.

Epstein, M. A., Shaywitz, S. E., & Shaywitz, B. A. & Woolston, J. L. (1991). The boundaries of attention deficit disorder. Journal of Learning Disabilities, 24, 78-86.

Faroane, S. V., Biederman, J., Chen, W. J., Milberger, S., Warburton, R., & Tsuang, M. T. (1995). Genetic heterogeneity in attention-deficit hyperactivity disorder (ADHD) : Gender, psychiatric comorbidity, and maternal ADHD. Journal of Abnormal Psychology, 104, 334-345.

Faraone, S. V., Biederman, J., Keenan, K., & Tsuang, M.T. (1991). A family-genetic study of girls with DSM-III attention deficit disorder. American Journal of Psychiatry, 148 (1), 112 - 117.

Gannon, L., Luchetta, T., Rhodes, K., Pardie, L., & Segrist, D. (1992). Sex bias in psychological research: Progress or complacency? American Psychologist, 47, 389-396.

Garb, H. N. (1997). Race bias, social class bias, and gender bias in clinical judgment. Clinical Psychology: Science and Practice, 4 (2), 99-120.

Garfinkel, B. D., & Amrami, K. K. (1992). A perspective on the attention-deficit disorders. Hospital and Community Psychiatry, 43 (5), 445-448.

Gaub, M., & Carlson, C.L. (1997). Gender differences in ADHD: a meta-analysis and critical review. Journal of the American Academy of Child and Adolescent Psychiatry, 36 (8), 1036 - 1045.

Gaultieri, T. & Hicks, R. E. (1985). An immuno-reactive theory of selective male affliction. The Behavioural and Brain Sciences, 8, 427-441.

Geshwing, N. & Galaburda, A. M. (1985). Cerebral lateralization. Biological mechanisms, associations and pathology. I-III. A hypothesis and program for research . Archives of Neurology, 42, 429-624.

Giedd, J. N., Snell, J. W., Lange, N., Rajapsakse, J. C., Kaysen, D., Vaituzis, A. C., Vauss, Y., Hamburger, S. D., Kozuch, P. L., & Rapoport, J. L. (1996). Quantitative magnetic resonance imaging of human brain development: Ages 4-18. Cerebral Cortex, 6, 551-560.

Glow, R. A. (1981). Cross validity and normative data on the Conners' parent and teacher rating scales. In K.D. Goodow and J. Loney (Eds.), The psychosocial aspects of drug treatment for hyperactivity (pp. 107-150). Boulder CO: Westview Press.

Golden, G.S. (1992). Commentary: The myth of attention-deficit hyperactivity disorder. Journal of Child Neurology, 7, 446-449.

Good, G. E., & Wood, P. K. (1995). Male gender role conflict, depression and help seeking: Do college men face double jeopardy? Journal of Counseling and Development, 74, 70-75.

Goodman, S. H., & Kohlsdorf, B. (1994). The development and psychopathology of conduct problems: Gender issues. In D. C. Fowles, P. Sutker, & S. H. Goodman (Eds.), Progress in experimental personality and psychopathology research. (Vol. 15, pp. 121-161). New York: Springer.

Goodman, R., & Stevenson, J. (1989). A twin study of hyperactivity: II. The aetiological role of genes, family relationships, and perinatal adversity. Journal of Child Psychology and Psychiatry, 30, 691-709.

Gordon, M. (1983). The Gordon Diagnostic System. Boulder, CO: Clinical Diagnostic Systems.

Gualtieri, T., & Hicks, R. (1985). An emmuno-reactive theory of selective male affliction. Behavioral and Brain Sciences, 8, 427-431.

Guilford, J. P. (1954). Psychometric methods (2nd ed.). New York: McGraw Hill.

Halperin, J. M., Matier, K., Bedi, G., Sharma, V., & Newcorn, J. (1992). Specificity of inattention, impulsivity, and hyperactivity to the diagnosis of attention deficit hyperactivity disorder. Journal of the American Academy of Child and Adolescent Psychiatry, 31 (2), 190-196.

Hamilton, S., Rothbart, M., & Dawes, R. M. (1986). Sex bias, diagnosis, and the DSM-III-R. Sex Roles, 15 (5/6), 269-274.

Hartung, C. M., & Widiger, T. A. (1998). Gender differences in the diagnosis of mental disorders: conclusions and controversies of The DSM-IV. Psychological Bulletin, 123 (3), 260-278.

Henker, B., & Whalen, C. K. (1989). Hyperactivity and attention deficits. American Psychologist, 44, 216 - 223.

Hinshaw, S.P. (1987). On the distinction between attentional deficits / hyperactivity and conduct problems / aggression in child psychopathology. Psychological Bulletin, 101 (3), 443 - 463.

Horn, W. F., Wagner, A. E., & Lalongo, N. (1989). Sex differences in school-aged children with pervasive attention deficit hyperactivity disorders. Journal of Abnormal Child Psychology, 17 (1), 109-125.

Huselid, R. F., & Cooper, M. L. (1994). Gender roles as mediators of sex differences in expressions of pathology. Journal of Abnormal Psychology, 103, 595-603.

Ingresoll, B. (1988). Your hyperactive child. New York: Doubleday.

James, A., & Taylor, E. (1990). Sex differences in the hyperkinetic syndrome of childhood. Journal of Child Psychology and Psychiatry, 31 (3) 437-446.

Jastak, S., & Wilkinson, G. S. (1984). The Wide-Range Achievement Test-Revised. Wilmington, DE: Jastak Associates.

Kagan, J. (1966). Reflection-impulsivity: The generality and dynamics of conceptual tempo. Journal of Abnormal Psychology, 71, 17-24.

Kaufman, A. S., & Kaufman, N. L. (1983). Kaufman Assessment Battery for Children: Administration and scoring manual. Circle Pines, MN: American Guidance service.

Kashini, J. H., Beck, N. C., Hoeper, E. W., Fallahi, C., Corcoran, C. M., McAllister, J. A., Rosenberg, T. K., & Reid, J. C. (1987). Psychiatric disorders in a community sample of adolescents. American Journal of Psychiatry, 144, 548, 589.

Kashani, J., Chapel, J. L., Ellis, J., & Shekim, W. O. (1979). Hyperactive girls. Journal of Operational Psychiatry, 10, 145-148.

Keenan, K., & Shaw, D. (1997). Developmental and social influences on young girls' early problem behavior. Psychological Bulletin, 121,95-113.

Kent, D. F., & O' Leary, K, D. (1976). A controlled evaluation of behavior modification with conduct problem children. Journal of Consulting and Clinical Psychology, 44, 586-596.

Kohn, A. (1989). Suffer the restless children. Atlantic Monthly, 264, 90-100.

Koriath, U., Gualtieri, C. T., Van Bourgondien, M. E., Quade, D., & Werry, J. S. (1985). Construct validity of clinical diagnosis in pediatric psychiatry: Relationship among measures. Journal of the American Academy of Child and Adolescent Psychiatry, 24, 429-436.

Lahey, B. B., Applegate, A., McBurnett, K., Biederman, J., Greenhill, L. L., Hynd, G. W., Barkley, R. A., Newcorn, J. A., Jensen, P., Richters, J. E., Garginkel, B., Kerdyk, L., Frick, P. J., Ollendick, T., Perex, D., Hart, E. L., Waldman, I., & Shaffer, D. (1994). DSM-IV field trials for attention deficit hyperactivity disorder in children and adolescents. American Journal of Psychiatry, 151 (11), 1673-1685.

Lambert, N. M., Sandoval, J., & Sassone, D. (1978). Prevalence of hyperactivity in elementary school as a function of social system definers. American Journal of Orthopsychiatry, 43, 446-463.

Laufer, M., Denhoff, E., & Solomons, G. (1957). Hyperkinetic impulse disorder in children's behavior problems. Psychosomatic Medicine, 19, 38-49.

Levine, M. D. (1992). Commentary: Attentional disorders: Elusive entities and their mistaken identities. Journal of Child Neurology, 7, 449-453.

Loney, J., & Milich, R. (1982). Hyperactivity, inattention, and aggression in clinical practice. Advances in Developmental and Behavioral Pediatrics, 3, 113-147.

Loring, M., & Powell, B. (1988). Gender, Race and DSM-III: A study of the objectivity of psychiatric diagnostic behavior. Journal of Health and Social Behavior, 29, 1-22.

Mannuzza, S., & Gittelman, R. (1984). The adolescent outcome of hyperactive girls. Psychiatry Research, 13, 19-29.

McGee, R., & Feehan, M. ( 1991). Are girls with problems of attention underrecognized? Journal of Psychopathology and Behavioral Assessment, 13 (3), 187 - 198.

McGee, R., Feehan, M., Williams, S., Partridge, F., Silva, P. A., & Kelly, J. (1990). DSM-III disorders in a large sample of adolescents. Journal of the American Academy of Child and Adolescent Psychiatry, 29, 611-619.

Mc Gee, R., Williams, S., & Silva, P. A. (1987). Comparison of girls and boys with teacher - identified problems of attention. Journal of the American Academy of Child and Adolescent Psychiatry, 26 (5), 711 - 717.

Meehl, P. E. (1967). Theory-testing in psychology and physics: A methodological paradox. Philosophy of Science, 34, 103-115.

Miller, R. G., Palkes, H. S., & Stewart, M. A. (1973). Hyperactive children in suburban elementary schools. Child Psychiatry and Human Development, 4, 121-127.

Mehrabian, A. (1971). Silent messages. Belmont, CA: Wadsworth.

Morgan, A. E., Hynd, G. W. Riccio, C. A., & Hall J. (1996). Validity of DSM-IV ADHD predominately inattentive and combined types: Relationship to previous DSM diagnoses / subtype differences. Journal of the American Academy of Child and Adolescent Psychiatry, 35 (3), 325-333.

Morrison, J. R., & Stewart, M. A. (1973). The psychiatric status of legal families of adopted hyperactive children. Archives of General Psychiatry, 28, 888-891.

Mulhern, S., Dworkin, P.H., Bernstein, B. (1994). Do parental concerns predict a diagnosis of attention-deficit hyperactivity disorder? Developmental and Behavioral Pediatrics, 15 (5), 384-352.

Myers, D. G. (1999). Social Psychology (6th ed). Boston: McGraw-Hill College.

Nichols, P. L., & Chen, T. C. (1980). Minimum brain dysfunction: A prospective study. Hillsdale, NJ: Erlbaum.

Ounsted, C. (1972). Gender & inheritance growth. In C. Ounsted & D. C. Taylor (Eds.), Gender Differences: Their ontology and significance (pp. 177-201). London: Churchill Livingstone.

Pauls. D. L., Shaywitz, S. E., Kramer, P. L., Shaywitz, B. A., & Cohen, D. J. (1983). Demonstration of vertical transmission of attention deficit disorder. Annals of Neurology, 14, 363.

Pekkanen, J. (2000). Making Sense of Ritalin. Readers Digest, 153-158.

Pelham, W. E., Evens, S.W., Gnagy, E. M., & Greenslade, K. E. (1992). Teacher ratings of DSM-III-R symptoms for the disruptive behavior disorders: prevalence, factor analyses, and conditional probabilities in a special education sample. School Psychology Review, 21 (2), 285 - 299.

Pelham, W. E., Gnagy, E. M., Greenslade, K. E., & Milich, R. ( 1992). Teacher ratings of DSM-III-R symptoms for the disruptive behavior disorders. Journal of the American Academy of Child and Adolescent Psychiatry, 31 (2), 210 - 218.

Pelham, W. E., Walker, J. L., Sturges, J., Hoza, J. (1989). Comparative effects of methylphenidate on ADD girls and ADD boys. Journal of the American Academy of Child and Adolescent Psychiatry, 28, 773-776.

Piel Cook, E., Warnke, M., & Dupuy, P. (1993). Gender bias and the DSM-III-R. Counselor Education and Supervision, 32, 311-322.

Popper, C. W., & Steingard, R. J. (1994). Disorders usually first diagnosed in infancy, childhood or adolescent. AIn R. E. Hales, S. C. Yudofsky, & J. A. Talbott (Eds.), Textbook Psychiatry (2nd ed., pp. 729-831). Washington: American Psychiatric Press.

Porteus, S. D. (1965). Porteus Maze Test: Fifty years' application. Palo Alto, CA: Pacific Books.

Prinz, R. J., Connor, P. A., & Wilson, C. C. ( 1981). Hyperactive and aggressive behaviors in childhood: Intertwined dimensions. Journal of Abnormal Child Psychology, 9 (2), 191 - 202.

Quay, H. C. (1983). A dimensional approach to behavior disorders: The Revised Problem Behavior Checklist. School Psychology Review, 12 (5), 244-249.

Rey, J. M. (1993). Oppositional defiant disorder. American Journal of Psychiatry, 150 (12), 1769 - 1778.

Rhee, S. H., Waldman, I. D., Hay, D. A., & Levy, F. (1999). Sex differences in genetic and environmental influences on DSM-III-R attention-deficit disorder. Journal of Abnormal Psychology, 108 (1), 24-41.

Robbins, L. N. (1991). Conduct disorder. Journal of Child Psychiatry and Psychology, 32, 193-212.

Ross, D. M., & Ross, S. A. (1982). Hyperactivity, current issues, research, and theory. New York: Wiley.

Rosvold, H. E., Mirsky, A. F., Sarason, I., Bransome, E. D., & Beck, L. H. (1956). A continuous performance test of brain damage. Journal of Consulting and Clinical Psychology, 20, 343-350.

Rutter, M. (1970). Sex differences in children's response to family stress. In E. J. Anthony & C. Koupernik (Eds.) The child and his family (pp. 165-196). New York: Wiley.

Rutter, M., Tizard, J., & Whitmore, K. (1970). Education, health and behaviour. London: Longmans.

Sabatino, D. A. & Vance, H. B. (1994). Is the diagnosis of attention deficit hyperactivity disorder meaningful? Psychology in the Schools, 31, 188-196.

Safer, D. J., & Zito, J. M. (1999). Psychotropic Medication for ADHD. Mental retardation and developmental disabilities research reviews, 5 (3), 237-242.

Safer, D. J., Zito, J. M., & Fine, E. M. (1996). Increased methylphenidate usage for attention deficit disorder in the 1990's. Pediatrics, 98 (6), 1084-1088.

Schachar, R., Rutter, M., & Smith, A. (1981). The characteristics of situationally and pervasively hyperactive children: Implications for syndrome definitions. Journal of Child Psychology and Psychiatry, 21, 293-311.

Schachar, R., Sandberg, S. & Rutter (1986). Agreement between teachers' ratings and observations of hyperactivity, inattentiveness, and defiance. Journal of Abnormal Child Psychology, 14 (2), 331 - 345.

Schaughency, E. A., & Rothlind, J. (1991). Assessment and classification of attention deficit hyperactive disorders. School Psychology Review, 20, 187-202.

Schrag, P., & Divoky, D. (1975). The myth of the hyperactive child. New York: Pantheon.

Shapiro, S. K., & Garfinkel, B. D. (1986). The occurrence of behavior disorders in children: The interdependence of attention deficit disorder and conduct disorder. Journal of The American Academy of Child Psychiatry, 25, 809-819.

Sharp, W. S., Walter, J. M., Marsh, W. L., Ritchie, G. F., Hamburger, S. D., & Castellanos, F. X. (1999). ADHD in girls: Clinical comparability of a research sample. Journal of the American Academy of child and Adolescent Psychiatry, 38 (1), 40-7.

Shaywitz, S. E., Shaywitz, B. A., Fletcher, J. M., & Escobar, M. D. (1990). Prevalence of reading disability in boys and girls. Results of the Connecticut Longitudinal Study. Journal of the American Medical Association, 264, 998-1002.

Sherman, J. A. (1980). Therapist attitudes and sex role stereotyping . In A. M. Brodsky, & R. T. Hare-Mustin (Eds.), Women and psychotherapy (pp. 35-66). New York: Guilford.

Silver, L. B. (1992). Attention-deficit hyperactivity disorder: A clinical guide to diagnosis and treatment. Washington: American Psychiatric Press, Inc.

Silverthorn, P., Frick, P. J., Kuper, K., & Ott, J. (1996). Attention deficit hyperactivity disorder and sex: A test of two etiological models to explain the male predominance. Journal of Clinical Child Psychology, 25, 52-59.

Skekim, W. O., Kashani, J., Beck, N., Cantwell, D., Martin, J., Rosenberg, J., & Costello, A. (1985). The prevalence of attention deficit disorders in a rural midwestern community sample of nine year-old children. Journal of the American Academy of Child Psychiatry, 24, 765-770.

Stevens, G. (1980). Bias in attributions of positive and negative behavior in children by school psychologists, parents, and teachers. Perceptual and Motor Skills, 50, 1283 - 1290.

Stevens, J., Quittner, A. L., & Abikoff, H. (1998). Factors influencing elementary school teachers' ratings of ADHD and ODD behaviors. Journal of Clinical Child Psychology, 27 (4), 406 - 414.

Stevens-Long, J. (1973). The effect of behavioral context on some aspects of adult disciplinary practice and affect. Child Development, 44, 476-484.

Still, G.F. (1902). Some abnormal physical conditions in children. Lancet, 1, 1008-1012.

Sunder, T. R. (1992). Commentary: Attention deficit disorder: Reductio ad absurdum. Journal of Child Neurology, 7, 454-558.

Swanson, J. M., McBurnett, K., & Wigal, T. (1993). Effect of stimulant medication on children with ADD: A "review of reviews." Exceptional Children, 60, 154-162.

Szatmari, P., Offord, D. R., & Boyle, M. H. (1989). Ontario Child Health Study: Prevalence of attention deficit disorder with hyperactivity. Journal of Child Psychology and Psychiatry, 30 (2), 219-230.

Szasz, T. (1974). The myth of mental illness (rev. ed.). New York: Harper and Row.

Szasz, T. (1992). Ideology and insanity. Essays on the psychiatric dehumanization of man. Syracuse, New York: Syracuse University Press.

Tanner, J. (1978). Fetus into man: Physical growth from conception to maturity. Cambridge, MA: Harvard University Press.

Tant, J. L., & Douglas, V. I. (1982) Problem solving in hyperactive, normal and reading disabled boys. Journal of Abnormal Child Psychology, 10, 285-306.

Taylor, D. (1985). Developmental rate is the major differentiator between the sexes. Behavioral and Brain Sciences, 8, 459-460.

Taylor, E., Hepinstall, E., Sonuga-Burk, & Sandberg, S. (1998, January). Sex differences in the prevalence of hyperactivity. Paper presented at the annual scientific meeting of the Royal College of Psychiatrists, London.

Taylor, D., & Ounsted, C. (1972). The nature of gender differences explored through ontogenetic analyses of sex ratios in disease. In C. Ounsted & D. Taylor (Eds.), Gender differences: Their ontogeny and significance (pp. 215-240). London: Churchill Livingstone.

Trites, R. L., Blouin, A. G., & Laprade, K. (1980). Factor analysis of the Conners' Teacher Rating Scale based on a large normative sample. Journal of Consulting and Clinical Psychology, 48, 615-621.

Trites, R. L., Dugas, E., Lynch, G., & Ferguson, H. B. (1979). Prevalence of hyperactivity. Journal of Pediatric Psychology, 4, 179-188.

Tsai, L., & Beisler, J. (1983). The development of sex differences in infantile autism. British Journal of Psychiatry, 142, 373-378.

Ullman, R. K., Sleator, E. K., & Sprague, R. L. (1985). Introduction to the use of the ACTeRS. Psychopharmacology Bulletin, 21, 915-916.

Walker, L. E. A. (1994). Are personality disorders gender biased? In S. A. Kirk & S. D. Einbinder (Eds.), Controversial issues in mental health (pp. 22-29). New York: Allyn & Bacon.

Wechsler, D. (1967). Manual for the Wechsler Preschool and Primary Scale of Intelligence. San Antonio: The Psychological Corporation.

Wechsler, D. (1974). Manual for the Wechsler Intelligence Scale for Children- Revised. San Antonio: The Psychological Corporation.

Weinberg, W. A., & Brumback, R. A. (1992). The myth of attention deficit disorder: Symptoms resulting from multiple causes. Journal of Child Neurology, 7, 431-445.

Werner, E., Bierman, J. M., French, F. E., Simonian, K., Connor, A., Smith, R. S., & Campbell, M. (1968). Reproductive and environmental casualties: a report of the 10-year followup of the children of the Kauai pregnancy study. Pediatrics, 42, 112-127.

Werry, J. S., Elkind, G. S., & Reeves, J. C. (1987). Attention deficit, conduct, oppositional, and anxiety disorders in children.III. Laboratory differences. Journal of Abnormal Child Psychology, 15, 409-428.

Widiger, T. A., & Spitzer, R. L. (1991). Sex bias in the diagnosis of personality disorders: Conceptual and methodological issues. Clinical Psychology Review, 11, 1-22.

Wirt, R. D., Lachar, D., Klinedinst, J. D., & Seat, P. D. (1984). Multidimensional description of child personality: A manual for the personality Inventory for Children-Revised 1984. Los Angeles: Western Psychological Services.

Wolraich, M. L., & Baumgaertel, A. (1997). The practical aspects of diagnosing and managing children with attention deficit hyperactivity disorder. Clinical Pediatrics, 497-504.

Wolraich, M. L., Hannah, J. N. Pinnock, T. Y., Baumgaertel, A., & Brown, J. (1996). Comparison of diagnostic criteria for attention-deficit hyperactivity disorder in a county-wide sample. Journal of the American Academy of Child and Adolescent Psychiatry, 35 (3), 319-324.

Zoccollilo, M. (1993). Gender and the development of conduct disorders. Development and Psychopathology, 5, 65-78.