

Teachers' Ratings of Disruptive Behaviors: The Influence of Halo Effects

**Howard Abikoff,^{1,3} Mary Courtney,¹ William E. Pelham, Jr.,² and
Harold S. Koplewicz¹**

This study evaluated the accuracy of teachers' ratings and examined whether these ratings are influenced by halo effects. One hundred thirty-nine elementary school teachers viewed videotapes of what they believed were children in regular fourth-grade classrooms. In fact, the children were actors who followed prepared scripts that depicted a child engaging in behaviors characteristic of an attention-deficit hyperactivity disorder (ADHD), an oppositional defiant disorder or a normal youngster. The findings provide support for a bias that was unidirectional in nature. Specifically, teachers rated hyperactive behaviors accurately when the child behaved like an ADHD youngster. However, ratings of hyperactivity and of ADHD symptomatic behaviors were spuriously inflated when behaviors associated with oppositional defiant disorder occurred. In contrast, teachers rated oppositional and conduct problem behaviors accurately, regardless of the presence of hyperactive behaviors. The implications of these findings regarding diagnostic practices and rating scale formats are discussed.

Teacher ratings are used extensively as diagnostic entry criteria and as outcome measures in clinical research with the disruptive behavior dis-

Manuscript received in final from March 8, 1993. Part of this research was presented at the Annual Meeting of the Society for Research in Child and Adolescent Psychopathology, Zandvoort, Holland, June 1991.

Preparation of this article was supported in part by a grant from the President's Fund of Long Island Jewish Medical Center awarded to Howard Abikoff.

The authors would like to express their thanks to Debra Murphy, Mary Hamilton, Karen Greenslade, and Diane Martin for their assistance in the early stages of the project.

¹Division of Child and Adolescent Psychiatry, Schneider Children's Hospital of Long Island Jewish Medical Center, New Hyde Park, New York 11042.

²Western Psychiatric Institute and Clinic, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15213.

³Address all correspondence concerning this article to Howard Abikoff, Ph.D., Division of Child and Adolescent Psychiatry, Schneider Children's Hospital, Long Island Jewish Medical Center, New Hyde Park, New York 11042.

orders (DBD). Teachers ratings are also relied on, often exclusively, to determine the prevalence rates of disorders in population samples. Teachers are valued as informants because they have the opportunity to observe children for relatively long periods of time in multiple schools settings. Moreover, their extensive experience with large numbers of youngsters provides teachers with an implicit normative data base against which to judge disordered behavior.

Although there are reports of the utility of teacher ratings in differentiating children with behavior problems (Atkins, Pelham, & Licht, 1989; Milich & Fitzgerald, 1985; Milich & Landau, 1988), there is also evidence of a lack of correspondence between teacher ratings and objective classroom observations. For example, Blunden, Spring, and Greenberg (1974) reported that children who behaved impulsively were rated by teachers as showing other behavior problems (e.g., restlessness, poor concentration, poor sociability) that were not substantiated by direct observation. Vincent, Williams, Harris, and Duval (1981) reported relatively poor agreement between teachers' ratings and direct observations of normal and hyperactive children. They suggested that the lack of correspondence was related to the influence of "negative halo effects" on teachers' ratings. Additional evidence that classroom ratings may be susceptible to halo effects is provided by Abikoff and Gittelman (1985a), who found that the average intercorrelation of observers' ratings of children's classroom behavior on an 11 item rating scale was .79, whereas the intercorrelation was only .24 for 11 observation codes scored by the same observers on the same children. Especially relevant is the report of a unidirectional bias in teacher ratings (Schachar, Sandberg, & Rutter, 1986). Specifically, children who behaved negatively toward others were rated by teachers as showing spuriously high levels of hyperactivity and inattentiveness. In contrast, teachers' ratings of children's behavior problems were accurate and were not influenced by observed levels of classroom hyperactivity. Such a unidirectional bias may underlie Prinz, Connor, and Wilson's (1981) finding that teachers more often recorded hyperactive behaviors on days when they reported aggressive behavior, whereas the conditional-probability of aggressive behaviors was not associated with the reported occurrence of hyperactive behaviors.

Population studies that rely on teachers' ratings for "diagnostic" classification typically report the high co-occurrence of attention-deficit hyperactivity disorder (ADHD) in children identified as having a conduct disorder (CD) or oppositional defiant disorder (ODD), relatively high rates of ADHD alone, and quite low rates of children identified as "pure" CD or ODD (e.g., Pelham, Gnagy, Greenslade, & Milich, 1992). These findings may depict the actual covariation of ADHD and other disruptive behaviors. On the other hand, in light of the reports of inaccuracies and biases in

teachers' ratings noted above, consideration needs to be given to the possible influence of halo effects on the classification patterns obtained in population studies.

In addition to these considerations, biases in teacher ratings may be a function of the nature of the factor items in commonly used teacher rating scales. For example, the Hyperactivity factor and Hyperkinesis Index of the Conners Teacher Rating Scale (Goyette, Conners, & Ulrich, 1978) include not only activity level and inattention items, but conduct problem items as well. It may be that inaccuracies in teachers' ratings are less common with the use of scales such as the IOWA Conners (Loney & Milich, 1982), which are not confounded by item content.

In summary, several different studies have reported inaccuracies in teacher ratings, related perhaps to the influence of unidirectional halo effects and rating scale characteristics, although few attempts have been made to investigate these issues directly. The aim of the current study was to manipulate systematically the actual classroom behaviors viewed by teachers so as to evaluate the accuracy of their ratings and examine whether and in what manner halo effects impact on teachers' ratings.

METHOD

Procedure

Regular and special education teachers viewed videotapes of children and were given the following instructions prior to the presentations of the tapes:

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.

Teachers were told not to open the questionnaire packet until the tape was completed. As expected, teachers asked if they should extrapolate from what was observed if the questions included behaviors that were not evident on the tape. They were instructed as follows:

This is 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.

Unbeknownst to the teachers, the children were actors who followed prepared scripts that varied as to the type and frequency of inappropriate child behaviors. The 10-min tapes depicted a boy engaging in behaviors characteristic of a "pure" ADHD, "pure" ODD, or normal youngster.⁴ As described below, the "pure" tapes were created so as to ensure as much as possible that (1) the ADHD child did not engage in oppositional behavior and (2) the rate of ADHD behaviors displayed by the ODD child was not in the deviant range and was equivalent to the rate of ADHD behaviors shown by the normal child.

The teachers were assigned to one of the two groups, with each group viewing, in counterbalanced order, the normal tape and either the ADHD or ODD tape. After each tape the teachers rated the target child on a 73-item 4-point scale (*Not at all, Just a little, Pretty much, Very much*) that included the following: (1) items comprising the Conners Teacher Rating Scale (CTRS) Hyperactivity factor and Hyperkinesis Index, (2) the IOWA Conners Aggression factor, which consists exclusively of items that assess oppositional defiant behaviors, (3) the IOWA Inattentive/Overactive factor, a putative index of hyperactive behaviors whose item content, unlike the Hyperactivity factor and Hyperkinesis Index, does not consist of conduct problem behaviors, (4) verbatim descriptors of the symptoms that compromise the ADHD and ODD DSM-III-R diagnostic categories, and (5) four DSM-III-R conduct disorder symptoms ("initiates physical fights," "steals," "lies," and "physically cruel").

Videotape Features

To address the study aims, the videotapes were developed so as to meet three goals. First, they had to be realistic; second, they had to depict behaviors that in type and frequency were characteristic of ADHD, ODD, and normal children and clearly differentiated that target children on "diagnostically" salient behaviors; third, to maximize the likelihood that teachers' ratings would be based on the behavior of the target child rather than

⁴It should be noted that our original intent was to include two additional tapes—one depicting a child with a "pure" conduct disorder and another of a child who was both ADHD and CD—and to examine the accuracy of teachers' ratings of these tapes as well. However, our independent assessments of the CD tape indicated that the child actor inadvertently engaged in excessive motor activity, thereby invalidating the utility of this tape. Similarly, the ADHD + CD tape that we produced did not have the appropriate behavior rates needed for the study. Consequently, only the data from the normal, ADHD, and ODD tapes are presented here.

on the behavior of others, there had to be control as well over the videotaped behaviors of the teacher and classmates toward the targeted child.

Various procedures were implemented to meet these objectives. Using child actors and an adult who served as the teacher, the tapes were made in an actual classroom and showed an academic lesson that required children to follow teacher instructions and to work independently in their seats. Prior to the study, the tapes were shown to an independent group of elementary school teachers, who wrote narratives about and rated the tapes as to the realism of the target child's behavior (e.g., "Have you had children like this in your classroom?") and the classroom setting (e.g., "What did the classroom seem like to you?"). The pilot data indicated that the teachers considered the child behavior and the setting to be realistic.

The behaviors engaged in by the target children were intended to capture those shown by ADHD, ODD, and normal children in the classroom. To this end we relied on clinical as well as empirical observations of clinic-referred and normal children. For the ADHD, ODD, and normal child tapes, we chose relative behavior rates based on classroom observation data collected previously on such children using two observations codes that have been found to differentiate ADHD and normal children: the revised Stony Brook observation code (Abikoff, Gittelman-Klein, & Klein, 1977; Abikoff, Gittelman, & Klein, 1980), and the Classroom Observations of Conduct and Attention Deficit Disorders (COCADD; Atkins, Pelham, & Licht, 1985, 1989).

Table I. Behavior Rates for Each Tape^a

Behavior	Normal	ADHD	ODD
Hyperactivity composite ^b	24	54	25
Solicitation	1	3	3
Noncompliance	0	1	4
Physical aggression	0	0	2
Verbal aggression to children	0	0	2
Verbal aggression to teacher	0	0	4

^aBased on Revised Stony Brook Classroom Observation Code (Abikoff & Gittleman, 1985b). ADHD = attention-deficit hyperactivity disorder; ODD = oppositional defiant disorder.

^bThe sum of the code categories: interference, off-task, gross-motor movements, out-of-chair, and minor-motor movements.

Ratings of the tapes by observers who employed the code's modified 15-sec time sampling procedure and operational definitions (see Abikoff & Gittelman, 1985b), and who were blind to the type of child depicted on the tape, indicated that our manipulations were successful. Thus, as can be seen in Table I, the Hyperactivity Composite rate for the ADHD child was twice that of the normal and ODD child. As intended, the rate of hyperactive behaviors shown by the child on the ODD tape was equivalent to that of the normal tape, whereas the negative behaviors characteristic of oppositional defiant disorder were coded only for the tape of the ODD child.

As a further check on the validity of the tapes, they were rated by an independent set of observers on another classroom observation coding system (Carlson, Pelham, Milich, & Dixon, 1992), a modified version of the COCADD (Atkins *et al.*, 1985, 1989). Here too, the tapes differed in the expected direction. (Percent of 20 sec intervals during which hyperactive behaviors occurred: ADHD tape = 50%, ODD = 29%, normal = 10%; percent of intervals during which oppositional/defiant behaviors occurred: ADHD = 7%, ODD = 35%, normal = 0%.)

To ensure that teachers' ratings were not influenced by nonspecific characteristics of the target child, the ADHD and ODD child were depicted by the same boy. The "normal" tape was played by another child and was rated by all the teachers. Finally, the classmates and teacher of the targeted children all followed scripts that controlled for the amount and type of interactions they had with the target, and with each other.

Subjects

The sample consisted of 139 elementary school teachers (72 regular education and 67 special education teachers) in the New York City public schools. One hundred thirty-three (96%) of the teachers were female. Seventy-six percent were Caucasian, 15% Hispanic, 6% African-American, and 3% other. They had considerable teaching experience, ranging from a mean of 12.7 years for regular education teachers to 15.6 years for those in special education. Seventy-five percent of the regular education teachers and 86% of those who taught in special education had a M.A. degree.

The teachers viewed the tapes in schools, and the number of teachers who attended the "screenings" ranged from 4 to 30. Seventy-nine teachers (41 regular education and 38 special education) viewed the normal/ADHD tapes and 60 (31 regular education and 29 special education) viewed the

Table II. Regular and Special Education Teachers' Mean Conners and IOWA Scale Factor Ratings of Each Tape^a

Factor	Type of child depicted on tape					
	Normal		ADHD		ODD	
	Reg	Spec	Reg	Spec	Reg	Spec
CTRS						
HA	0.34 (0.28)	0.36 (0.34)	2.17 (0.42)	1.90 (0.61)	2.06 (0.49)	1.73 (0.59)
HI	0.54 (0.37)	0.53 (0.42)	1.91 (0.40)	1.66 (0.55)	2.14 (0.42)	1.80 (0.45)
IOWA						
I/O	3.44 (2.44)	3.73 (2.73)	10.83 (2.20)	10.05 (2.90)	10.27 (0.29)	9.28 (2.67)
A	0.44 (1.16)	0.22 (0.67)	4.59 (3.70)	4.42 (4.54)	12.10 (2.28)	10.62 (2.69)

^aNote: Numbers in parentheses are standard deviations. Clinical cut-scores for HA and HI = 1.5, I/O = 9, A = 6. For both the regular (Reg) and special education (Spec) teachers, the ratings of the normal tape combine those teachers who viewed the ADHD or ODD tapes, since their ratings did not differ significantly. Because of missing ratings, N = 137 for HA, and N = 138 for HI, I/O, and A. CTRS = Conners Teacher Rating Scale; HA = Hyperactivity factor; HI = Hyperkinesis Index; I/O = Inattentive/Overactive factor; ADHD = Attention-Deficit Hyperactivity Disorder; ODD = Oppositional Defiant Disorder.

normal/ODD tapes. The teachers each received \$30.00 for their participation.

RESULTS

Data Analysis

Two (Type of Tape: ADHD, ODD) × two (Type of Teacher: Regular, Special) ANCOVAs were carried out on the teachers' ratings for each of the teacher scale factors. The teachers' ratings of the "normal" tape were considered representative of their individual response styles and therefore served as covariates. Logistic regression analyses were conducted on the percentage of teachers in each group who rated the target as meeting DSM-III-R criteria for the various disruptive disorders. ("Meeting criteria" required a rating of at least *pretty much* on the requisite number of DSM-III-R symptom items, i.e., 8 for ADHD, 5 for ODD, and 3 for CD.)

Table III. Percentage of Teachers Whose Ratings Met DSM-III-R Diagnostic Criteria^a

Type of child depicted on tape	Percentage whose ratings met diagnostic criteria for:		
	ADHD	ODD	ADHD+ODD
ADHD	63.3	16.7	14.1
ODD	40.0	90.0	40.0
Normal	2.2	0.0	0.0

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

Teacher Effects

Special and regular education teachers did not differ significantly ($p > .50$) in their ratings of oppositional behaviors, nor were any significant interactions (all $p > .20$) found between type of tape and type of teacher. However, on the ADHD and ODD tapes, the regular education teachers rated the targets as significantly higher on the Hyperactivity factor, $F(1, 135) = 11.87, p < .001$, Hyperkineses Index, $F(1, 135) = 13.51, p < .001$, and Inattentive/Overactive factor $F(1, 135) = 5.20, p < .03$ (see Table II).

ADHD Judgments

Factor Ratings. The mean CTRS Hyperactivity factor scores for the ADHD and ODD tapes were *not* significantly different, and *both* were above the clinical/research cut-score of 1.5. Moreover, as can be seen in Table II, not only were the ADHD and ODD tapes rated above the cut-score on the Hyperkineses Index, but the overall mean Hyperkineses Index rating of the ODD tape ($M = 2.14$) was significantly *higher* than the rating of the ADHD tape ($M = 1.91$), $F(1, 135) = 7.32, p < .008$. Notably, a nonsignificant difference was obtained for ratings on the IOWA Inattentive/Overactive factor. Here too, mean scores for both tapes were above the clinical cut-score on the Inattentive/Overactive factor (Pelham, Milich, Murphy, & Murphy, 1989).

DSM-III-R Ratings. As noted in Table III, the ratings of 63% of the teachers classified the child in the ADHD tape as meeting DSM-III-R criteria for ADHD, whereas the tape depicting an ODD child was rated

as meeting ADHD criteria by 40% of the teachers; a nonsignificant difference, $\chi^2(1, N = 139) = 0.80$. In contrast, as expected, the normal tape was rated as meeting diagnostic criteria for ADHD by very few (2.2%) of the teachers.

ODD Judgments

Factor Ratings. The ODD tape was rated significantly higher than the ADHD tape, $F(1, 135) = 176.92, p < .001$, on the IOWA Aggression factor. Moreover, the mean Aggression rating for the ODD tape (Table II) was well above the clinical cut-score of 6 for this factor (Pelham *et al.*, 1989). In contrast, though the mean Aggression ratings of the ADHD tape were significantly higher than the Aggression ratings given to the normal tape, $t(78) = 9.83, p < .001$, they were below the clinical cut-score.

DSM-III-R Ratings. The ratings of 90 percent of the teachers who viewed the child on the ODD tape met criteria for ODD. In comparison, for the child on the ADHD tape, the ratings of significantly fewer teachers (16.7%) met DSM-III-R criteria for $\chi^2(1, N = 139) = 22.19, p < .001$. None of the teachers rated the child on the normal tape as ODD.

ADHD + ODD Judgments

The ratings of 40% of the teachers who viewed the ODD tape met diagnostic criteria for ADHD and ODD. This "dual-diagnosis" rating for the ADHD tape was obtained from significantly fewer teachers (14%), $\chi^2(1, N = 139) = 5.32, p < .05$. None of the teachers' ratings of the normal tape met the dual diagnosis of ADHD and ODD.

CD Judgments

The ratings from only three (3.8%) teachers who viewed the ADHD tape and two (3.3%) who saw the ODD tape resulted in the child meeting DSM-III-R criteria for CD. These rates did not differ significantly. No teachers' ratings of the normal tape met CD criteria. The teachers' overall accuracy in not rating the target children as conduct disorder is reflective of the general absence of conduct disorder behaviors displayed on the tapes. As noted in Table I, other than the two instances of physical aggression shown by the ODD child, no other conduct disorder symptoms were depicted by any of the children.

DISCUSSION

The findings provide support for a bias in teacher ratings, but one that is only unidirectional in nature. Both regular and special education teachers tended to rate ADHD behaviors with a fair degree of accuracy when a child behaved like an ADHD youngster. However, teacher factor ratings of hyperactivity and of ADHD symptomatic behaviors were spuriously inflated when a child engaged in problematic behaviors associated with oppositional defiant disorder. In contrast, teachers rated oppositional and conduct problem behaviors accurately, regardless of whether or not a child also demonstrated hyperactive behaviors. As we discuss below, these findings suggest that caution needs to be exercised in using teacher ratings as the sole criteria to characterize and subdivide ADHD and ODD children.

The Stony Brook Classroom Observation Code, which served as the primary check of the intended experimental manipulations, indicated that the rate of hyperactive behaviors depicted on the ODD tape was a *non-deviant* rate that was almost identical to the rate of hyperactive behaviors shown on the normal tape. Yet, whereas the child on the ODD tape received ADHD and hyperactivity ratings from the teachers that were in the clinical range, the ratings of the child in the normal tape were in the normal range. This clearly suggests that the elevated ADHD ratings of the ODD child were a result of the oppositional/defiant behaviors displayed by this child.

A second observation system (the COCADD), whose observation codes and definitions were somewhat different from the Stony Brook system, was used as another criterion for external validity. Scores on the COCADD codes indicated that the rates of hyperactive behaviors on the ODD tape were approximately midway between the rates of hyperactive behaviors coded for the ADHD and normal tapes. If the teachers' ratings were unbiased, then their hyperactive ratings of the ODD tape should have been about 40% *lower* than the ratings given to the ADHD tape, reflecting the difference in the relative rates of COCADD-scored hyperactive behaviors on the two tapes. In fact, the teachers' hyperactivity ratings of the two tapes were equivalent, strongly suggesting that the findings were related to the influence of halo effects and not solely to possible differences in ADHD behaviors in the tapes.

In considering the generalizability of the results, certain aspects of this analogue study design need to be kept in mind. First, the videotapes, though presented as representing genuine classroom behaviors, were simulated and each consisted of a 10-min sample of behavior. In practice, teachers' ratings are usually based on overall judgments of behaviors that occur

over time frames lasting days or weeks, and we did not measure these time frames. Although it is unclear to what degree teachers' ratings in practice are susceptible to the kind of halo effects found in this study, it is parsimonious to assume that they reflect similar processes. For example, Schachar *et al.* (1986) found that teacher ratings covering a 2-week period of actual classroom behavior result in the same type of unidirectional bias reported here. Clearly additional research is needed to address potential biases in natural settings.

It has been noted that children who are rated as conduct disordered/aggressive by teachers are more likely to also be rated hyperactive than vice versa (Hinshaw, 1987). However, this asymmetry is most often found in studies that use the Conners Teacher Rating Scale factors, within which ADHD item content is confounded by conduct problem behaviors (e.g., "disturbs other children" is on the Hyperactivity factor and "temper outbursts and unpredictable behavior" is on the Hyperkinesis Index). Consequently, Hinshaw emphasized the need for purer screening measures when assessing the degree of overlap between ADHD and CD/ODD. Yet, the current study found the same asymmetrical pattern, not only on the CTRS factors, but also on the IOWA Conners, which consists of scales that were empirically developed to minimize the relationship between hyperactivity and conduct problem behaviors. Given that these findings were obtained in a setting that experimentally controlled the actual behaviors observed by the teachers, consideration needs to be given to the possible influence of unidirectional halo effects on the asymmetrical classification of "comorbidity" patterns found when using teacher scales. This bias could help explain the somewhat surprising and unexpected finding by Pelham, Evans, Gnagy, and Greenslade (1992) and Pelham, Gnagy, *et al.* (1992) that teachers' ratings of DSM-III-R ODD behaviors of boys in both special and regular education classes predicted the presence of ADHD as well as did the teachers' ratings of ADHD behaviors.

It is conceivable that the unidirectional bias reported here resulted from the influence of teachers' implicit personality theories regarding disruptive children. For example, given the high comorbidity rates common among the three disruptive disorders, perhaps the teachers' ratings reflect their experiences or assumptions regarding the co-occurrence of disruptive behaviors, rather than the actual behaviors depicted on the tapes. However, the findings argue against this notion. Specifically, although ADHD children are often oppositional and many have an ODD diagnosis, the child on the ADHD tape was not rated ODD. Moreover, even though CD behaviors are extremely prevalent in children with ODD, only 2% of the teachers rated the child on the ODD tape as meeting criteria for conduct disorder, indicating that problematic behaviors that did not occur were not

reported by the teachers. Therefore, the most plausible interpretation of the findings is that oppositional behavior creates a negative halo resulting in elevated ADHD ratings. It seems highly unlikely that teachers' implicit personality theories would operate specifically on their hyperactivity judgments about ODD children without influencing any other behavior ratings.

The results of the current study have other implications. They support the notion that teacher ratings should not be relied on exclusively for the classification or diagnosis of ADHD. Rather, such ratings should serve as only one of multiple sources of relevant information (see Lahey *et al.*, 1987; Loeber & Lahey, 1988). Moreover, the findings suggest that if a youngster receives elevated teacher ratings of conduct problems *and* hyperactivity, be it from regular or special education teachers, more specific information should be obtained from the teacher to clarify and validate that ADHD behaviors are actually occurring in school at the level reported. The inclusion of semistructured teacher interviews, like those used with parents that allow for probing and clarification of responses, might be helpful in these cases, as would the use of classroom observation procedures presumably less influenced by halo effects (e.g., Abikoff *et al.*, 1977; Atkins *et al.*, 1989).

The absence of any interaction effects between tape and type of teacher indicate that unidirectional halo effects operate in both regular and special education teachers. However, when judging the same samples of behavior, regular education teachers give higher ratings of hyperactivity than do special education teachers. This difference may reflect a lower rating threshold for ADHD behaviors among regular education teachers, related perhaps to their less frequent exposure to these behaviors compared to their special education counterparts. Further, this difference between teachers also suggests that developers of teacher rating scales must consider the type of teacher when generating behavioral norms for such scales, since the mean cut-scores may be different for children in regular and special education classes. For example, these data would suggest that the cut-scores for children in special education would be a full point lower than for children in regular education classes on the IOWA Conners Inattentive/Overactive factor (cf. Pelham *et al.*, 1989).

Several research questions also stem from the study findings. First, it is important to determine whether a similar unidirectional bias influences *parent* ratings of hyperactive behaviors. This is especially relevant in light of a recent report that suggests that a halo effect may influence clinician ratings of ADHD symptoms in children with conduct problems (Newcorn, Sharma, Matier, Hall, & Halperin, 1992). Second, it would be useful to determine whether other established teacher rating scales are less prone to rater bias effects than the CTRS and IOWA Conners. Third, further work is needed to determine whether the inclusion of more specific op-

erational definitions of rating scale items can reduce rater bias effects (Siegel, Dragovich, & Marholin, 1976) and improve diagnostic precision (Sandoval, 1981). Similarly, perhaps the use of rating scale items that require descriptive rather than evaluative judgments might reduce the influence of negative halo effects (Mintz & Collins, 1985). Fourth, consideration needs to be given to the development and evaluation of programs that attempt to train teachers to improve the accuracy of their ratings. Such efforts might focus not only on teaching systematic observational procedures (see Weinrott, 1977), but also utilize models of clinical and person perception expertise that focus on the detection of patterns of covariation between antecedent events and target behaviors (Dawson, Zeitz, & Wright, 1989).

Finally, studies of the factors underlying the unidirectional bias are warranted. Several possible mechanisms may be operating. For example, it may be simply that during at least some instances of oppositional behavior, an ODD child is inattentive, in motion, and/or impulsive, as a necessary consequence of exhibiting the ODD behaviors (e.g., the defiant verbalization has been an impulsive interruption). Alternatively, aggressive and disruptive behaviors are extremely salient, unacceptable, and the most aversive of deviant classroom behaviors for teachers (Walker, Reavis, Rhode, & Jensen, 1985). When these aversive behaviors are paired with ADHD behaviors, teachers may be sensitized to this pairing because of their association in a Pavlovian sense (Prinz *et al.*, 1981).

Moreover, does the occurrence of *particular* problem behaviors increase the likelihood of a negative halo and elevated ADHD ratings? For example, if defiance directed toward adults is particularly salient and distressing for teachers, the arousal and stress caused by the child's defiant behavior may cloud teachers' perceptions, making them especially susceptible to the factors that give rise to biases in ratings. Recent laboratory studies have shown that children exhibiting ADHD/ODD behaviors substantially increase stress in caretaking adults, but no study has teased out which among these externalizing behaviors produce the largest impact in adults (Pelham & Lang, in press). Relatedly, Mintz and Collins (1985), in an analogue study, reported that college students' ratings of the motor activity of videotaped adults were inflated when the adult behaved in an inappropriate and conspicuous manner. The authors speculated that teachers may be affected by a similar negative halo in response to salient child behaviors. Further, is it likely that ADHD behaviors will be rated inaccurately if aversive and inappropriate behaviors occur in an impulsive manner (Blunden *et al.*, 1974)? Investigations of teachers' ratings of videotapes wherein distinct behaviors are experimentally manipulated might help to clarify these issues.

REFERENCES

- Abikoff, H., & Gittelman, R. (1985a). The normalizing effects of methylphenidate on the classroom behavior of ADHD children. *Journal of Abnormal Child Psychology*, *13*, 33-44.
- Abikoff, H., & Gittelman, R. (1985b). Classroom observation code: A modification of the Stony Brook code. *Psychopharmacology Bulletin*, *21*, 901-909.
- Abikoff, H., Gittelman, R., & Klein, D. F. (1980). A classroom observation code for hyperactive children: A replication of validity. *Journal of Consulting and Clinical Psychology*, *48*, 555-565.
- Abikoff, H., Gittelman-Klein, R., & Klein, D. F. (1977). Validation of a classroom observation code for hyperactive children. *Journal of Consulting and Clinical Psychology*, *45*, 772-783.
- 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*, 155-167.
- 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*, 423-435.
- Blunden, D., Spring, C., & Greenberg, L. M. (1974). Validation of the Classroom Behavior Inventory. *Journal of Consulting and Clinical Psychology*, *42*, 84-88.
- 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*, 213-232.
- Dawson, V. L., Zeitz, C. M., & Wright, J. C. (1989). Expert-novice differences in person perception: Evidence of experts' sensitivities to the organization of behavior. *Social Cognition*, *7*, 1-30.
- Goyette, C. H., Conners, C. K., & Ulrich, R. F. (1978). Normative data on revised Conners parent and teacher rating scales. *Journal of Abnormal Child Psychology*, *6*, 221-236.
- Hinshaw, S. P. (1987). On the distinction between attentional deficits/hyperactivity and conduct problems/aggression in child psychopathology. *Psychological Bulletin*, *101*, 443-463.
- Lahey, B. B., McBurnett, K., Piacentini, J. C., Hartdagen, S., Walker, J., Frick, P. J., & Hynd, G. W. (1987). Agreement of parent and teacher rating scales with comprehensive clinical assessments of attention deficit disorder with hyperactivity. *Journal of Psychopathology and Behavioral Assessment*, *9*, 429-440.
- Loeber, R. & Lahey, B. B. (1988). Recommendations for research on disruptive behavior disorders of childhood and adolescence. In B. B. Lahey & A. E. Kazdin (Eds.), *Advances in clinical child psychology* (Vol. 12, pp. 221-251). New York: Plenum Press.
- Loney, J., & Milich R. (1982). Hyperactivity, inattention, and aggression in clinical practice. In M. Wolraich & D. K. Routh (Eds.), *Advances in developmental and behavioral pediatrics* (Vol. 3, pp. 113-147). Greenwich, CT: JAI Press.
- Milich, R., & Fitzgerald, G. (1985). Validation of inattention/overactivity and aggression ratings with classroom observations. *Journal of Consulting and Clinical Psychology*, *53*, 139-140.
- Milich, R., & Landau, S. (1988). Teacher ratings of inattention/overactivity and aggression: Cross-validation with classroom observations. *Journal of Clinical Child Psychology*, *17*, 92-97.
- Mintz, L. I., & Collins, B. (1985). Qualitative influences on the perception of movement: An experimental study. *Journal of Abnormal Child Psychology*, *13*, 143-153.
- Newcorn, J. H., Sharma, V., Matier, K., Hall, S., & Halperin, J. M. (1992, February). *Comorbidity of ADHD: Differential effect of oppositional/conduct and anxiety disorders on individual symptom domains*. Poster presented at the annual meeting of the Society for Research in Child and Adolescent Psychopathology, Sarasota, FL.
- Pelham, W. E., Jr., Evans, 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, 285-299.
- Pelham, W. E., Jr., 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, 210-218.
- Pelham, W. E., & Lang, A. R. (in press). Parental alcohol consumption and deviant child behavior: Laboratory studies of reciprocal effects. *Clinical Psychology Review*.
- Pelham, W. E., Milich, R., Murphy, D. A., & Murphy, H. A. (1989). Normative data on the IOWA Conners teacher rating scale. *Journal of Clinical Child Psychology*, 18, 259-262.
- Prinz, R. J., Conner, P. A., & Wilson, C. C. (1981). Hyperactive and aggressive behaviors in childhood: Intertwined dimensions. *Journal of Abnormal Child Psychology*, 9, 191-202.
- Sandoval, J. (1981). Format effects in two teacher rating scales of hyperactivity. *Journal of Abnormal Child Psychology*, 9, 203-218.
- Schachar, R., Sandberg, S., & Rutter, M. (1986). Agreement between teachers' ratings and observations of hyperactivity, inattentiveness, and defiance. *Journal of Abnormal Child Psychology*, 14, 331-345.
- Siegel, L. J., Dragovich, S. L., & Marholin, D., II (1976). The effects of biasing information on behavioral observations and rating scales. *Journal of Abnormal Child Psychology*, 4, 221-233.
- Vincent, J. P., Williams, B., Harris, G. E., Jr., & Duval, G. C. (1981). Classroom observations of hyperactive children: A multiple validation study. In K. D. Gadow & J. Loney (Eds.), *Psychosocial aspects of drug treatment for hyperactivity* (pp. 207-248). Boulder, CO: Westview Press.
- Walker, H. M., Reavis, H. K., Rhode, G., & Jenson, W. R. (1985). A conceptual model for delivery of behavioral services in behavior disordered children in educational settings. In P. H. Bornstein & A. E. Kazdin (Eds.), *Handbook of clinical behavior therapy with children* (pp. 700-741). Homewood IL: Dorsey Press.
- Weinrott, M. R. (1977). Improving the validity of global ratings. *Journal of Abnormal Child Psychology* 5, 187-197.