J Psychopathol Behav Assess (2010) 32:128–137
DOI 10.1007/s10862-009-9135-3

Halo Effects in Ratings of ADHD and ODD: Identification of Susceptible Symptoms

Cynthia M. Hartung · Elizabeth K. Lefler · Ashley B. Tempel · Monica L. Armendariz · Benjamin A. Sigel · Carolyn S. Little

Published online: 14 April 2009

Abstract Halo effects in the assessment of ADHD and ODD were examined. Participants were 159 undergraduate college students who rated children described as showing disruptive behaviors. Bidirectional halo effects were found. Specifically, the presence of oppositionality artificially inflated ratings of inattention and hyperactivity, and the combined presence of inattention and hyperactivity artificially inflated ratings of oppositionality. Several specific items were found to be particularly susceptible to halo effects. Due to these halo effects caution should be exercised when diagnosing multiple behavior disorders, especially with items found to be particularly susceptible. Clinical interviews conducted by mental health professionals may help distinguish between the true presence of multiple disorders and halo effects based on ratings. Future research should determine whether structured interviews conducted by mental health professionals are less susceptible to halo effects than rating scales.

Keywords Halo effects · Disruptive behaviors · Assessment · Rating scales

C. M. Hartung (✉) · C. S. Little
Department of Psychology, University of Wyoming,
1000 East University Avenue, Department #3415,
Laramie, WY 82071, USA
e-mail: chartung@uwyo.edu

E. K. Lefler · M. L. Armendariz · B. A. Sigel
Department of Psychology, Oklahoma State University,
Stillwater, OK, USA

A. B. Tempel
Department of Psychology, West Virginia University,
Morgantown, WV, USA

Two disruptive behavior disorders (DBDs) in the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (*DSM-IV*; American Psychiatric Association (APA), 2000) are Attention-Deficit/Hyperactivity Disorder (ADHD) and Oppositional Defiant Disorder (ODD). Halo effects have been documented in ratings of ADHD and ODD such that a child exhibiting one behavior (e.g., oppositionality) is rated as having other behaviors (e.g., inattention and hyperactivity) without any direct evidence of these other behaviors (Abikoff et al. 1993; Schachar et al. 1986; Stevens et al. 1998; Jackson and King 2004; Hartung et al. 2006).

Halo effects have been referred to as *unidirectional* or *bidirectional* (Jackson and King 2004; Hartung et al. 2006). Unidirectional halo effects occur when a child exhibiting a certain behavior (e.g., oppositionality) is rated as showing other behaviors (e.g., inattention and hyperactivity) but the effect only occurs in one direction. Bidirectional halo effects occur when a child exhibiting a certain behavior (e.g., oppositionality) is rated as showing other behaviors (e.g., inattention and hyperactivity) and a child displaying the second set of behaviors (e.g., inattention and hyperactivity) is also rated as exhibiting the first behavior (e.g., oppositionality). Halo effects might result in inappropriate diagnosis of a second disruptive behavior disorder (DBD) following appropriate diagnosis of one DBD.

This issue is further complicated because the research suggests that there is a high rate of comorbidity between ADHD and ODD, which has been estimated at 45 to 84% (Barkley 2006). Furthermore, the DBD symptom dimensions (i.e., inattention, hyperactivity and oppositionality) are highly correlated with one another (e.g., r=.58 to .77; Burns et al. 2001). Therefore, although there are high levels of comorbidity among children with DBDs, this fact does not fully explain the halo effects found in the aforemen-

tioned research studies. There are at least two reasons to believe that comorbidity alone cannot fully explain halo effects. First, when raters are asked to rate a child's behavior it is presumed that they will only endorse symptoms that have been directly observed. Thus, if raters endorse symptoms that they have not observed, this cannot be considered comorbidity because the child was not actually displaying two discrete sets of behaviors. Second, it is possible that some of the supposed comorbidity between these two disorders could actually be a result of halo effects. That is, the high correlations among inattention, hyperactivity, and oppositionality may be partially due to halo effects and true comorbidity rates may be lower than the research suggests. Clearly more research is needed regarding halo effects, comorbidity, and DBDs.

It is also important to understand halo effects in the assessment of ADHD and ODD because these disorders are typically assessed with parent and teacher ratings and interviews (McMahon and Frick 2005; Pelham et al. 2005). Studies have shown that ratings from teachers and college students are susceptible to unidirectional (Abikoff et al. 1993; Stevens et al. 1998) and bidirectional (Hartung et al. 2006; Jackson and King 2004; Schachar et al. 1986) halo effects. Unidirectional halo effects have been found using videotaped vignettes and teacher raters (Abikoff et al. 1993; Stevens et al. 1998). In addition, bidirectional halo effects have been found using live observations and teacher raters (Schachar et al. 1986), videotaped vignettes and teacher raters (i.e., Jackson and King 2004), and written vignettes and college student raters (Hartung et al. 2006).

Some of the details of these studies will be reviewed because they are important for understanding the degree to which the results can be compared across studies. In the Schachar et al. (1986) study, teachers rated 1st grade boys in their classrooms using the Conners Teacher Rating Scale (CTRS; Conners 1969). Also, research assistants, who were blind to the CTRS scores, conducted live observations of these boys for a total of two weeks. The boys were rated for 10 minute blocks of time every hour during the school day. Teachers also completed the CTRS at the end of each observation week. Results showed high agreement between teacher CTRS ratings and ratings of observed behaviors. However, both the teacher ratings and the live observations were prone to bidirectional halo effects. Specifically, hyperactivity ratings were artificially inflated by the presence of defiance and inattention, inattention ratings were artificially inflated by the presence of defiance, and behavior problem ratings, including defiance, were artificially inflated by the presence of inattention and hyperactivity.

In the Abikoff et al. (1993) study, teachers watched videotaped vignettes of 4th grade male child actors in two of three conditions. Each teacher watched and rated a typical boy and either a boy displaying *DSM-III-R* ADHD symptoms or *DSM-III-R* ODD symptoms. Ratings included the verbatim ADHD and ODD criteria. The results showed that teachers were fairly accurate in their ADHD ratings of the boy displaying ADHD symptoms and in their ODD ratings of the boy displaying ODD symptoms. However, teacher ratings were susceptible to unidirectional halo effects such that their ratings of hyperactivity were inflated for the boy displaying ODD symptoms. In the Stevens et al. (1998) study, teachers watched the Abikoff et al. (1993) videotapes of male child actors and rated the boys using the CTRS and the SNAP-IV (Swanson and Carlson 1994). The results were similar in that teacher ratings of hyperactivity were inflated for the boy displaying ODD symptoms, showing evidence of a unidirectional halo effect. The results using the SNAP-IV showed a slightly less severe halo effect and the authors concluded that this was due to the SNAP-IV being more structured than the CTRS. These authors also found that background knowledge of ADHD and exposure to educational materials did not improve accuracy or decrease this halo effect.

More recently, Jackson and King (2004) created new videotapes of 4th grade male and female actors using the transcripts from the Abikoff et al. (1993) study. Again, children with ADHD or ODD were portrayed in addition to typical children. Jackson and King asked teachers to complete the CTRS and the DBD rating scale (Pelham et al. 1992) after viewing a videotaped vignette. The DBD rating scale is based on the *DSM-III-R* criteria for inattention, hyperactivity and oppositionality. Results showed bidirectional halo effects for boys and girls. Specifically, the presence of oppositionality artificially inflated ratings of inattention and hyperactivity. In addition, the presence of inattention and hyperactivity artificially inflated ratings of oppositionality.

Finally, Hartung et al. (2006) created vignettes depicting 3rd grade boys and girls displaying ADHD-Combined Type (ADHD-CT), ADHD-Inattentive Type (ADHD-IT), ODD and depression. Hartung et al. (2006) used the depression vignette as a control condition. College students were asked to read each vignette while listening to an audio-taped reading of the vignette. Next, they were asked to complete a *DSM-IV*-based rating scale for the child described in each vignette. Similar to the results from Schachar et al. (1986) and Jackson and King (2004), bidirectional halo effects were found. Specifically, the presence of oppositionality (i.e., ODD) artificially inflated ratings of inattention and hyperactivity and the presence of inattention and hyperactivity (i.e., ADHD-CT) artificially inflated ratings of oppositionality. Consistent with the Jackson and King study, these results held regardless of the sex of the child in the vignette. Hartung et al. also used an ADHD-IT vignette and found that the presence of inattention (i.e.,

ADHD-IT) artificially inflated ratings of hyperactivity. This finding is consistent with the results of the live observation study (i.e., Schachar et al. 1986).

Taken together, these studies suggest that halo effects may contribute to misdiagnosis if rating scales are the primary assessment tool. Additional research is needed to determine if: (1) parent and/or mental health professional ratings will produce halo effects, (2) unidirectional or bidirectional halo effects will be replicated in future studies, and (3) specific symptoms are particularly susceptible to halo effects.

The current study examined whether specific symptoms are particularly susceptible to halo effects using written vignettes and college student raters. None of the studies that have identified halo effects in ratings of ADHD and ODD (Abikoff et al. 1993; Schachar et al. 1986; Stevens et al. 1998; Jackson and King 2004; Hartung et al. 2006), have examined whether specific symptoms are particularly susceptible to halo effects. If certain symptoms are identified as particularly susceptible, clinicians could be warned to take special care with these items when considering differential or multiple diagnoses. Thus, the goal of current study was to determine whether particular ADHD and ODD symptoms are susceptible to halo effects.

As in the Hartung et al. (2006) study, bidirectional halo effects were expected. Specifically, it was predicted that a child displaying oppositionality (i.e., ODD) would be rated as inattentive and hyperactive and that a child displaying inattention and hyperactivity (i.e., ADHD-CT) would be rated as oppositional. Furthermore, it was expected that a child displaying inattention (i.e., ADHD-IT) would be rated as hyperactive. It was also hypothesized that ADHD and ODD items with conceptual overlap (i.e., the behavior could be due to ADHD or ODD) would be more susceptible to halo effects than items that are unique to one disorder. For example, *Avoids, dislikes or is reluctant to engage in work that requires sustained mental effort* could be endorsed due to inattention (cannot complete the task) or oppositionality (will not complete the task); whereas *Often actively defies or refuses to comply with adults' requests or rules* would presumably only be endorsed as a symptom of oppositionality.

Method

Participants One hundred fifty-nine undergraduate college students participated in the study (74 men, 85 women). Participants ranged in age from 18 to 35 years (M=20.39, SD=2.72). The racial/ethnic composition of the adult raters was 81.1% European American, 7.5% Native American, 5.0% African American, 2.5% Asian, 1.9% Hispanic and 1.3% other.

Vignettes Four vignettes describing the behavior of an individual child were used. In all of the vignettes, the child was described as being 8-years-old. The vignettes range in length from 231 words (i.e., Typical vignette) to 291 words (i.e., ADHD-CT vignette). Three vignettes were modified versions of those used in Hartung et al. (2006). In the ADHD-Inattentive Type (ADHD-IT) vignette, reference was made to five symptoms of inattention. In the ADHD-CT vignette, reference was made to five symptoms of inattention and five symptoms of hyperactivity. In the ODD vignette, reference was made to six symptoms of oppositionality. When *DSM-IV* symptoms were referenced, they were either verbatim or paraphrased. For example, in the ADHD-CT vignette, the *DSM-IV* symptom "often leaves seat" was paraphrased as "often gets out of her seat." Similarly, in the ODD vignette, the *DSM-IV* symptom "is often spiteful or vindictive" was paraphrased as "he often attempts to get even." Finally, a *Typical* vignette was created for this study. In the Typical vignette, a child was described as displaying typical, age-appropriate behavior for an 8-year-old and reference was not made to any symptoms of inattention, hyperactivity or oppositionality. There were male and female versions of each of the four vignettes; with the exception of changes in the child's first name and associated pronouns, the male and female versions were identical. Thus, there were a total of eight vignettes (i.e., four vignette types with male and female versions).

Rating scales The standardized 26-item Disruptive Behavior Rating Scale (DBRS; Barkley and Murphy 2006) was used. The DBRS consists of *DSM-IV* ODD symptoms (8 items) and ADHD symptoms (9 symptoms of inattention and 9 symptoms of hyperactivity). The DBRS uses a 4-point scale with options of *never/rarely, sometimes, often,* or *very often.*

Expert ratings As a test of validity, five licensed psychologists rated the vignettes. These psychologists had 8 to 10 years of post-doctoral clinical experience and all had considerable experience with children with DBDs. The experts rated the child in each vignette significantly higher on target variables than on non-target variables. Target variables were the symptom dimension(s) portrayed at high levels in the vignette (e.g., hyperactivity and inattention in the ADHD-CT vignette). Non-target dimensions were the symptom dimensions that were not portrayed in the vignette (e.g., hyperactivity in the ODD vignette).

Mean levels of symptom endorsement as well as average agreement were calculated based on expert ratings. Agreement among raters was calculated by averaging the percentage of expert raters who: (1) endorsed target symptoms that were referenced in the vignette, (2) failed

to endorse target symptoms that were not referenced in the vignette, and (3) failed to endorse non-target symptoms. Average agreement was calculated separately for target symptoms that were specifically referenced in the vignette and those that were not.

For the ADHD-IT vignette, experts rated the child as displaying a mean of 6.6 inattention symptoms. In contrast, experts rated the child in the ADHD-IT vignette as showing no oppositionality symptoms and 0.2 hyperactivity symptoms. For inattention symptoms that were specifically referenced in the vignette, average agreement that these symptoms were present was 92.0%. For inattention symptoms that were not specifically referenced in the vignette, average agreement that these symptoms were absent was 50.0%. For hyperactivity symptoms, none of which were referenced in the vignette, average agreement that these symptoms were absent was 93.3%. For oppositionality symptoms, none of which were referenced in the vignette, average agreement that these symptoms were absent was 97.5%.

For the ADHD-CT vignette, experts rated the child as showing 7.2 inattention symptoms, 7.0 hyperactivity symptoms, and no oppositionality symptoms. For inattention symptoms that were specifically referenced in the vignette, average agreement that these symptoms were present was 84.0%. For inattention symptoms that were not specifically referenced in the vignette, average agreement that these symptoms were absent was 25.0%. For hyperactivity symptoms that were specifically referenced in the vignette, average agreement that these symptoms were present was 88.0%. For hyperactivity symptoms that were not specifically referenced in the vignette, average agreement that these symptoms were absent was 30.0%. For oppositionality symptoms, none of which were referenced in the vignette, average agreement that these symptoms were absent was 100.0%.

For the ODD vignette, experts rated the child as showing 7.6 oppositionality symptoms, 1.4 inattention symptoms and 1.2 hyperactivity symptoms. For inattention symptoms, none of which were referenced in the vignette, average agreement that these symptoms were absent was 82.2%. For hyperactivity symptoms, none of which were referenced, average agreement that these symptoms were absent was 86.7%. For oppositionality symptoms that were specifically referenced in the vignette, average agreement that these symptoms were present was 93.3%. For oppositionality symptoms that were not specifically referenced, average agreement that these symptoms were absent was 0%. Thus, when oppositionality was a target symptom, expert raters endorsed ODD symptoms regardless of whether the symptoms were specifically referenced.

For the Typical vignette, experts rated the child as showing no inattention symptoms, 0.2 hyperactivity symptoms, and no oppositionality symptoms. For inattention symptoms, none of which were referenced in the vignette, average agreement that these symptoms were absent was 100.0%. For hyperactivity symptoms, none of which were referenced in the vignette, average agreement that these symptoms were absent was 97.8%. For oppositionality symptoms, none of which were referenced in the vignette, average agreement that these symptoms were absent was 100.0%.

These results provide evidence of the validity of the vignettes for differentiating among the symptom clusters in each vignette. All the vignettes produced mean symptom counts above the *DSM-IV* threshold on target variables and below the *DSM-IV* threshold on non-target variables. In addition, expert raters did not rate the child in the Typical vignette as showing significant symptoms of any DBD dimension. On target dimensions, average agreement among raters regarding the presence of symptoms that were specifically referenced in the vignettes was very high (i.e., 84% to 93%). However, on target dimensions average agreement among raters regarding the absence of symptoms that were not specifically referenced in the vignettes was variable (i.e., 0% to 50%). On non-target dimensions, average agreement among raters regarding the absence of symptoms was quite high (i.e., 82% to 100%).

Procedure Procedures for recruiting participants and study method were approved by the institutional review board at the university where the research was conducted and these procedures were in compliance with the ethical standards of the American Psychological Association (APA). The current study was completed via the internet. Each participant rated one vignette for each of the four behavior types, and was randomly assigned to rate two boys and two girls. Participants were unable to look back at the vignette while they completed the rating scale; they could not return to previously viewed screens once they had advanced. Presentation order was set to ensure that sex of the child in the vignette was alternated, and the order of vignette type was counterbalanced. For each vignette, participants were instructed to "Mark the box that best describes the behavior of this child" for each rating scale item. Finally, participants were provided with a written debriefing statement and were thanked for their participation.

Data analyses Symptom counts on each dimension (i.e., inattention, hyperactivity, and oppositionality) for each vignette (i.e., ADHD-IT, ADHD-CT, ODD, and Typical) served as DVs. Thus, there were a total of 12 DVs. Responses were converted by considering a symptom present if *often* or *very often* was endorsed and considering a symptom absent if *never/rarely* or *sometimes* was endorsed. The symptoms considered present on the dimension were summed to create symptom counts. For the

ADHD-IT and ODD vignettes, there was one target and two non-target variables. For the ADHD-CT vignette, there were two target variables and one non-target variable. For the Typical vignette, there were three non-target variables. Multiple statistical comparisons were conducted and family-wise Bonferroni alpha corrections were calculated.

The first hypothesis was that a child displaying oppositionality would be rated as inattentive and hyperactive, and that a child displaying ADHD-CT would be rated as oppositional. To test for these halo effects, a series of two-tailed paired samples *t*-tests were conducted, comparing symptom counts on variables from the DBD vignettes (i.e., ADHD-IT, ADHD-CT and ODD) to symptom counts on respective variables from the Typical vignette. A total of nine paired samples *t*-tests were conducted. A Bonferroni correction resulted in an alpha cutoff of $p<.006$ (.05/9=.006).

The second hypothesis was that ADHD and ODD items with conceptual overlap (i.e., behavior could be due to ADHD or ODD) would be more susceptible to halo effects than items that are unique to one disorder. In order to determine whether individual ADHD and ODD items were particularly susceptible to halo effects, additional comparisons were conducted. For non-target item dimensions that resulted in significantly higher item counts in a DBD vignette than in the Typical vignette, two sets of comparisons were conducted to identify susceptible items.

First, the rate of endorsement for each item on the non-target dimension in the DBD vignette was compared to the rate of endorsement for the respective item on the Typical vignette (e.g., inattention items in the ODD vignette were compared to inattention items in the Typical vignette). This series of comparisons was conducted to determine if items were more likely to be endorsed as non-target items in a DBD vignette than as non-target items in the Typical vignette. Thus, if an item was endorsed significantly more often in a non-target DBD vignette than in a non-target Typical vignette preliminary evidence of susceptibility was assumed. McNemar's test was used to compare pairs of proportions for these comparisons (Sheskin 2007).

Second, the rate of endorsement for each item on the non-target dimension in the DBD vignette was compared to the rate of endorsement for the respective item in the DBD vignette(s) where the dimension was targeted (e.g., inattention items from the ODD vignette were compared to inattention items from the ADHD-CT and ADHD-IT vignettes). It was expected that target DBD items would be endorsed significantly more often than non-target DBD items. Thus, if a non-target DBD item was endorsed as often as, or more often than, a target DBD item, this was considered additional evidence of susceptibility. Again, McNemar's test was used to compare pairs of proportions for these comparisons (Sheskin 2007). For each DBD dimension, 2 or 3 sets of 9 paired comparisons were conducted. Thus, a Bonferroni correction resulted in an alpha value of $p<.006$ (.05/9=.009).

Finally, as a measure of effect size, binominal confidence intervals were calculated for each item identified as being particularly susceptible to halo effects. Binomial confidence intervals were calculated for each susceptible item based on the first set of comparisons but not the second. The first set of comparisons involved determining if the probability of endorsement of a non-target item in a DBD vignette was significantly greater than the probability of endorsement of the respective non-target item in the Typical vignette. However, the second set of comparisons involved determining if the probability of endorsement of a non-target item in a DBD vignette was similar to, or greater than, the probability of endorsement of the respective target item in a DBD vignette. Therefore, the binomial confidence interval was not calculated as a measure of effect size for this set of comparisons.

Results

Statistical significance of halo effects For the ADHD-IT vignette, the inattention symptom count (a target dimension) was expected to be significantly higher than the inattention symptom count in the Typical vignette. This comparison was statistically significant (see Table 1) suggesting that the vignette effectively portrayed ADHD-IT. Next, the hyperactivity symptom count (a non-target symptom) was expected to be significantly higher than in the Typical vignette, showing a halo effect of ADHD-IT to hyperactivity. This halo effect was statistically significant and resulted in a medium effect size (see Fig. 1). Finally, the oppositionality symptom count (a non-target symptom) was not significantly higher than in the typical vignette, showing no halo effect for ADHD-IT to oppositionality.

For the ADHD-CT vignette, the inattention symptom count (a target symptom) was expected to be significantly higher than in the Typical vignette. In addition, the hyperactivity symptom count (a target symptom) was expected to be significantly higher than in the Typical vignette. Both comparisons were statistically significant (see Table 1) suggesting that the vignette effectively portrayed ADHD-CT. Next, based on previous studies, the oppositionality symptom count (a non-target symptom) was expected to be significantly higher than oppositionality in the Typical vignette showing a halo effect from ADHD-CT to oppositionality. This halo effect was statistically significant with a medium effect size (see Fig. 1).

For the ODD vignette, the inattention symptom count (a non-target symptom) was expected to be significantly higher than in the Typical vignette showing a halo effect

Table 1 Paired comparisons between respective DBD and typical vignette dimensions

	DBD vignette		*Typical vignette*		*Paired comparisons*		
	M	*SD*	*M*	*SD*	*t-test*	*p-value*	*Cohen's d*
ADHD-IT							
Inattention[a]	5.78	1.96	0.36	0.96	31.96	.000	2.53[e]
Hyperactivity[b]	1.77	2.23	0.89	1.25	4.85	.000	0.38[c]
Oppositionality[b]	0.36	1.20	0.18	0.69	1.73	.043	0.14
ADHD-CT							
Inattention[a]	7.35	1.78	0.36	0.96	40.67	.000	3.22[e]
Hyperactivity[a]	6.99	1.82	0.89	1.25	33.80	.000	2.68[e]
Oppositionality[b]	1.64	2.37	0.18	0.69	7.99	.000	0.63[d]
ODD							
Inattention[b]	3.22	2.81	0.36	0.96	12.96	.000	1.03[e]
Hyperactivity[b]	3.94	2.62	0.89	1.25	14.73	.000	1.17[e]
Oppositionality[a]	7.29	1.43	0.18	0.69	57.85	.000	4.59[e]

For inattention and hyperactivity, possible scores ranged from 0 to 9. For oppositionality, possible scores ranged from 0 to 8. N=159 for all variables and df=158 for all t-tests

[a] Target dimensions are the behaviors that were described in the vignette

[b] Non-target dimensions are the behaviors that were not described in the vignette

[c] Small effect size (d>.20); [d] Medium effect size (d>.50); [e] Large effect size (d>.80)

from ODD to inattention. In addition, the hyperactivity symptom count (a non-target symptom) was expected to be significantly higher than in the Typical vignette showing a halo effect from ODD to hyperactivity. Both halo effects were statistically significant and resulted in large effects sizes (see Table 1 and Fig. 1). Finally, the oppositionality symptom count (a target symptom) was expected to be significantly higher than in the Typical vignette. This comparison was statistically significant suggesting that the vignette effectively portrayed ODD.

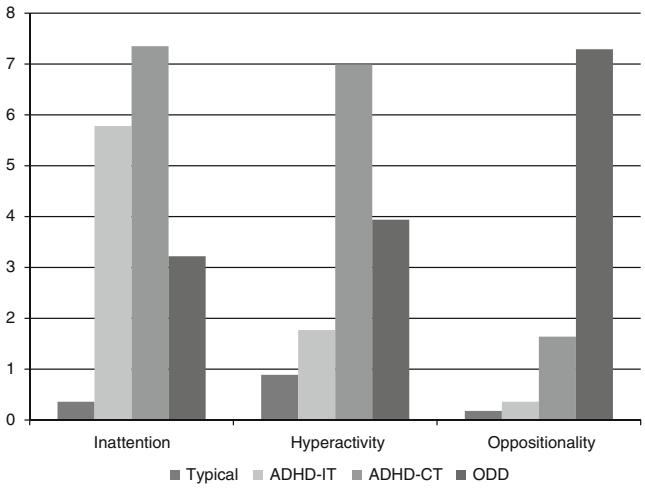

Fig. 1 Symptom counts for each DBD dimension by vignette

Item susceptibility to halo effects Given that four symptom count comparisons were significant for halo effects, individual items were examined to determine which items, if any, were particularly susceptible. First, non-target inattention items in the ODD vignette were examined. All inattention items in the ODD vignette were endorsed significantly more often than the respective items in the Typical vignette (see Table 2). In addition, none of the inattention items in the ODD vignette were endorsed as often as, or more than respective items in ADHD-CT vignette. However, two inattention items in the ODD vignette were endorsed as often as, or more often than, respective items in the ADHD-IT vignette. Thus, these two items (i.e., *Doesn't listen when spoken to directly; Avoids, dislikes or is reluctant to engage in work that requires sustained mental effort*) appeared particularly susceptible to this halo effect involving artificially inflated levels of inattention in the presence of ODD. The binomial confidence interval for the *Doesn't listen* item was 51.50 to 102.00 suggesting that this item was 52 to 102 times more likely to be endorsed in the ODD vignette than in the Typical vignette. The binomial confidence interval for the *Avoids* item was 3.17 to 5.30 suggesting that this item was 3 to 5 times more likely to be endorsed in the ODD than Typical vignette.

Next, halo effects involving hyperactivity items from the ODD vignette were examined. Eight out of nine non-target hyperactivity items in the ODD vignette were endorsed

Table 2 Percentage of raters who endorsed inattention items in the ODD vignette contributing to a halo effect

	ADHD-CT Target	ADHD-IT Target	ODD Non-target	Typical Non-target
1. Fails to give close attention to details or makes careless mistakes in his/her work[f]	82.2[a]	63.9[b]	28.9[c]	6.3[d]
2. Has difficulty sustaining his/her attention in tasks or fun activities[e, f]	91.0[a]	72.0[b]	23.9[c]	5.1[d]
3. Doesn't listen when spoken to directly[e]	81.8[a]	11.3[b]	65.0[c]	0.6[b]
4. Doesn't follow through on instructions and fails to finish work[e]	86.6[a]	70.1[a]	50.9[b]	5.0[c]
5. Has difficulty organizing tasks and activities[f]	76.6[a]	79.7[a]	31.0[b]	1.3[c]
6. Avoids, dislikes or is reluctant to engage in work that requires sustained mental effort	71.5[a]	47.1[b]	40.9[b]	7.5[c]
7. Loses things necessary for tasks and activities[e, f]	82.9[a]	68.4[a]	20.1[b]	0.6[c]
8. Is easily distracted[e, f]	98.1[a]	91.7[a]	36.1[b]	8.3[c]
9. Is forgetful in daily activities	71.5[a]	79.6[a]	26.6[b]	1.3[c]

Percentages for the same item without any common superscripts are significantly different based on McNemar's test ($p<.006$)

[e] Item was specifically referenced in the ADHD-CT vignette

[f] Item was specifically referenced in the ADHD-IT vignette

significantly more often than in the Typical vignette (see Table 3). Next, these eight non-target hyperactivity items were compared to the respective items in the ADHD-CT vignette. Three hyperactivity items in the ODD vignette (i.e., *Blurts out answers before questions have been completed; Has difficulty awaiting turn; Interrupts or intrudes on others*) were endorsed as often as, or more often than, the respective items in the ADHD-CT vignette. Consequently, these three hyperactivity items appeared particularly susceptible to this halo effect involving artificially inflated levels of hyperactivity in the presence of ODD. The binomial confidence interval for the *Blurts out* item was 4.92 to 8.83 suggesting that this item was 5 to 9 times more likely to be endorsed in the ODD vignette than in the Typical vignette. The binomial confidence interval for the *Awaiting turn* item was 4.67 to 8.83 suggesting that this item was also 5 to 9 times more likely to be endorsed in the ODD than Typical vignette. The binomial confidence interval for the *Interrupts* item was 10.29 to 15.57 suggesting that this item was 10 to 16 times more likely to be endorsed in the ODD than Typical vignette.

Next, halo effects involving hyperactivity items from the ADHD-IT vignette were examined. Five hyperactivity items in the ADHD-IT vignette were endorsed significantly more often than in the Typical vignette (see Table 3). However, none of these five hyperactivity items in the ADHD-IT vignette were endorsed as often as the respective items in the ADHD-CT vignette (see Table 3). As a result, none of the hyperactivity items appeared particularly susceptible to this halo effect involving artificially inflated levels of hyperactivity in the presence of ADHD-IT.

Finally, halo effects involving oppositionality items from the ADHD-CT vignette were examined. Seven out of eight non-target oppositionality items in the ADHD-CT vignette were endorsed significantly more often than in the Typical vignette (see Table 4). However, none of these seven oppositionality items in the ADHD-CT vignette were endorsed as often as, or more often than, respective items

Table 3 Percentage of raters who endorsed hyperactivity symptoms in the ADHD-IT and ODD vignettes contributing to halo effects

	ADHD-CT Target	ADHD-IT Non-target	ODD Non-target	Typical Non-target
1. Fidgets with hands or feet or squirms in seat[d]	93.0[a]	26.4[b]	20.1[b]	4.5[c]
2. Leaves seat in the classroom or in other situations in which remaining seated is expected[d]	96.9[a]	27.8[b]	34.0[b]	1.9[c]
3. Seems restless	89.2[a]	29.1[b]	38.2[b]	7.0[c]
4. Has difficulty engaging in leisure activities and doing fun things quietly	74.1[a]	15.2[b]	42.8[b]	4.4[c]
5. Seems "on the go" or acts as if "driven by a motor"[d]	94.8[a]	27.0[b]	34.0[b]	38.0[b]
6. Talks excessively[d]	77.7[a]	12.0[b]	43.0[c]	17.7[b]
7. Blurts out answers before questions have been completed	39.6[a]	8.9[b]	34.2[a]	3.8[b]
8. Has difficulty awaiting turn[d]	73.9[a]	18.4[b]	63.5[a]	7.7[c]
9. Interrupts or intrudes on others	66.0[a]	12.7[b]	87.1[c]	4.4[b]

Percentages for the same item without any common subscripts are significantly different based on McNemar's test ($p<.006$)

[d] Items was specifically referenced in the ADHD-CT vignette

Table 4 Percentage of raters who endorsed oppositionality symptoms in the ADHD-CT vignette contributing to a halo effect

	ADHD-CT Non-target	ADHD-IT Non-target	ODD Target	Typical Non-target
1. Loses temper[d]	22.4[a]	6.3[b]	96.8[c]	4.4[b]
2. Argues with adults[d]	23.1[a]	2.5[b]	98.1[c]	1.9[b]
3. Actively defies or refuses to comply with adult requests[d]	38.4[a]	5.8[b]	97.5[c]	1.3[b]
4. Deliberately annoys people[d]	22.2[a]	3.8[b]	91.1[c]	0.6[b]
5. Blames others for his/her mistakes or misbehavior[d]	13.5[a]	5.7[a]	92.4[b]	2.5[a]
6. Is touchy or easily annoyed by others	19.5[a]	7.5[b]	82.3[c]	3.8[b]
7. Is angry or resentful	13.9[a]	1.9[b]	89.3[c]	1.9[b]
8. Is spiteful or vindictive[d]	12.7[a]	2.5[b]	86.8[c]	1.9[b]

Percentages for the same item without any common subscripts are significantly different based on McNemar's test ($p<.006$)

[d] Item was specifically referenced in the ODD vignette

in the ODD vignette (see Table 4). Therefore, none of the oppositionality items were deemed particularly susceptible to this halo effect involving artificially inflated levels of oppositionality in the presence of ADHD-CT.

Clinical significance of halo effects The magnitude of the halo effects was examined to determine clinical significance. In the ODD vignette, the mean symptom count for inattention was 3.22 (*SD*=2.81) and for hyperactivity was 3.94 (*SD*=2.63). For the ADHD-CT vignette, the mean for oppositionality was 1.64 (*SD*=2.37). For the ADHD-IT vignette, the mean for hyperactivity was 1.77 (*SD*=2.23). These levels do not reach the *DSM-IV* cutoffs of six symptoms of inattention and/or hyperactivity for ADHD or four symptoms of oppositionality for ODD.

Discussion

This study examined ratings of ADHD and ODD to identify symptoms that may be particularly susceptible to halo effects. Bidirectional halo effects were found for ADHD-CT and ODD. Specifically, children displaying ODD symptoms were rated as having artificially inflated levels of inattention and hyperactivity/impulsivity symptoms. In addition, children displaying ADHD-CT symptoms were artificially rated as having symptoms of oppositionality. These findings are consistent with previous studies showing bidirectional halo effects (Hartung et al. 2006; Jackson and King 2004; Schachar et al. 1986). Including the current study, four studies to date have shown bidirectional halo effects and two have shown unidirectional halo effects (Abikoff et al. 1993; Stevens et al. 1998). Neither the format of presentation nor the raters (i.e., teachers or college students) predicted whether the effects were unidirectional or bidirectional; specifically, one of the bidirectional studies used teacher-rated videotaped vignettes and both of the unidirectional studies used teacher-rated videotaped vignettes. Future research is indicated to determine whether unidirectional or bidirectional effects predominate and whether there is an interaction between rater (i.e., teacher, parent, college student, or mental health professional) and directionality of effect (i.e., unidirectional or bidirectional).

In addition to the effects across ODD and ADHD, children displaying ADHD-IT symptoms were rated as having symptoms of hyperactivity. Thus, the presence of inattention (i.e., ADHD-IT) artificially inflated ratings of hyperactivity. This unidirectional effect within ADHD diagnoses was also found by Hartung et al. (2006) and Schachar et al. 1986). The videotape vignette studies did not examine this effect because an ADHD-IT vignette was not included (i.e., Abikoff et al. 1993; Jackson and King 2004; Stevens et al. 1998). In the current study, bidirectionality of this within-ADHD halo effect could not be tested because an ADHD-Predominately Hyperactive Type (ADHD-HT) vignette was not included. Thus, future research should include an ADHD-HT vignette.

This study also examined individual symptoms that were particularly susceptible to halo effects. There were two inattention symptoms that were frequently endorsed in the ODD vignette. First, *Doesn't listen when spoken to directly* was quite susceptible to the halo effect. One reason for this could be that the wording from the DBRS deviated from the *DSM-IV* which states, *Does not seem to listen when spoken to directly*. The deletion of the word *seem* may suggest that the child is actively ignoring the adult rather than not listening because he/she is focused on something else. This wording change might have biased this item to be more susceptible to this halo effect. When clinicians are using rating scales or interviews including this item, they should take special care to note the wording that is used.

The second inattention item that appeared particularly susceptible to halo effects in the ODD vignette was *Avoids, dislikes, or is reluctant to engage in work that requires sustained mental effort.* Since the wording of this item in the DBRS was not changed in a meaningful way from that

in the *DSM-IV* (i.e., the only difference was the use of *work* instead of *tasks*), the susceptibility of this item for halo effects may be due to conceptual overlap. That is, the rater must determine whether the child is avoiding work that requires sustained mental effort because the child has attention problems and *cannot* focus for long or because the child is oppositional and *will not* focus. It is possible that the addition of a qualifier in the future editions of the *DSM* (e.g., *not due to oppositional behavior*), might be helpful in decreasing the susceptibility of this item to halo effects. It is notable that the halo effect was not entirely accounted for by these two symptoms. Specifically, when these two symptoms were removed from the analyses, there was still a statistically significant halo effect ($p<.001$). Thus, although these two symptoms were particularly susceptible to this halo effect there appears to be a significant halo effect resulting from inattention being generally inflated in the ODD vignette.

There were also three hyperactivity-impulsivity symptoms frequently endorsed in the ODD vignettes (i.e., *Blurts out answers before questions have been completed; Has difficulty awaiting turn; Interrupts or intrudes on others*). The wording of these three symptoms in the DBRS is identical to that in the *DSM-IV* and all three items are included in the *DSM-IV* impulsivity subscale rather than the hyperactivity subscale. Again, the susceptibility of these items may be due to conceptual overlap. Children who are impulsive may have trouble waiting in line, blurting out answers, and interrupting simply because they are impulsive. In contrast, children who are oppositional may refuse to wait in line due to defiance and may blurt out answers and interrupt because they are prone to arguing with adults or have difficulty with anger. Again, adding qualifiers to these items in the future editions of the *DSM* may help decrease susceptibility. For example, the qualifier *not due to oppositionality* could be added to *Has difficulty awaiting turn.* Similarly, the qualifier *not due to argumentativeness* could be added to the interrupting and blurting out items. Again, the halo effect was not entirely accounted for by these three symptoms. Specifically, when these three symptoms were removed from the analyses, there was still a statistically significant halo effect ($p<.001$). Thus, although these three symptoms were particularly susceptible to this halo effect there appears to be a significant halo effect resulting from hyperactivity being generally inflated in the ODD vignette.

No items were identified as particularly susceptible for two additional statistically significant halo effects. First, there was a halo effect involving non-target hyperactivity in the ADHD-IT vignette. Second, there was a halo effect involving non-target oppositionality in the ADHD-CT vignette. As with the halo effects for inattention and hyperactivity in the ODD vignette, there appears to be significant halo effects resulting from hyperactivity being generally inflated in the ADHD-IT vignette and to oppositionality being generally inflated in the ODD vignette. Even though no individual items appeared particularly susceptible to these halo effects, these effects were still statistically significant.

Taken together, the current findings suggest that clinicians should use care when considering multiple DBD diagnoses because the halo effects between ADHD and ODD may result in a diagnosis of false comorbidity. In terms of clinical significance, the magnitude of the halo effects found in this study would likely not affect a child with a clinically significantly level of symptoms for one DBD (e.g., 6 symptoms of oppositionality) and no symptoms of another DBD (e.g., 0 symptoms of inattention). However, for a child with a clinically significant level of symptoms for one DBD (e.g., 6 symptoms of oppositionality) and some symptoms of another DBD (e.g., 3 symptoms of inattention) halo effects could increase ratings to a clinically significant level. Therefore, when a child presents with a mixed set of DBD symptoms, there is a risk that a comorbid DBD will be incorrectly diagnosed. Thus, clinicians should be particularly careful when: (1) a child with ODD is exhibiting a few symptoms of inattention and/or hyperactivity; (2) a child with ADHD-CT is exhibiting a few symptoms of ODD; and/or (3) a child with ADHD-IT is exhibiting a few symptoms of hyperactivity. In addition, clinicians should be particularly careful with the five particularly susceptible items identified in this study.

The identification of items that are particularly susceptible to halo effects supports a hypothesis proposed by Abikoff et al. (1993). These authors suggested that halo effects may result from the high degree of conceptual overlap among symptoms of inattention, hyperactivity and oppositionality. For example, a child described as oppositional might *have difficulty following through on instructions* even though this is a *DSM-IV* inattention symptom. Similarly, a child described as inattentive and hyperactive, might *refuse to comply with adult requests* even though this is a *DSM-IV* oppositionality symptom. Therefore, symptoms that require the rater to determine if a child *cannot* (due to ADHD) or *will not* (due to ODD) do what is asked may be particularly susceptible to halo effects. Some of the items identified in this study as particularly susceptible to halo effects are those that show conceptual overlap.

It is also of note that several target symptoms that were not specifically referenced in the vignettes were endorsed as present by the expert raters (e.g., within the trait of inattention *is often easily distracted* is conceptually similar to *often has difficulty sustaining attention*). It is possible that halo effects are also inflating ratings within target dimensions. Specifically, if several symptoms of inattention are observed, this may result in these inattention symptoms

being endorsed as well as additional inattention items. However, because the number of expert raters was so small, this finding could not be examined statistically. Future research should examine the possibility of halo effects within target dimensions, and the possible inflation of within trait symptom counts.

As mentioned previously, these halo effects suggest caution when diagnosing DBDs. In addition, clinical interviews conducted by mental health professionals with parents and teachers may help distinguish between the true presence of multiple disorders and apparent comorbidity based on ratings. Although the incremental validity of semi-structured or structured interviews, beyond that obtained from ratings, has been questioned (e.g., Pelham et al. 2005; Wolraich et al. 2003), the current findings suggest that mental health professionals might use interviews to confirm whether these symptoms were endorsed because the child's inability to complete certain tasks is related to oppositionality and defiance (i.e., ODD) or inattention or hyperactivity (i.e., ADHD). One reason that interviews may be less susceptible to halo effects than parent or teacher ratings is that mental health professionals may be less susceptible to halo effects than parents or teachers. However, it remains to be determined whether structured interviews conducted by mental health professionals are less susceptible to halo effects than rating scales. Future research in this area is warranted.

There are some notable limitations to the current study. First, this study was an analogue study and may not be representative of ratings completed by parents and teachers who know a child and observe behavior over months and years. Second, written vignettes and college student raters were used in this study. Thus, the generalizability of the current findings is limited. More research is needed, not only with parents and mental health professionals as raters, but also with ratings of actual children.

All of the halo effect studies support the recommendation to collect data from multiple informants when conducting DBD assessments (Abikoff et al. 1993; Hartung et al. 2006; Jackson and King 2004; Schachar et al. 1986; Stevens et al. 1998). Future research will be useful for determining whether these halo effects extend to parents and whether mental health professionals are less susceptible to halo effects. If halo effects extend to mental health professionals, then the use of semi-structured interviews may not be as helpful as anticipated. However, until this research is conducted, it is recommended that semi-structured interviews be used in combination with rating scales when considering multiple DBD diagnoses.

Acknowledgements We thank the following individuals for their contributions to this project: Chelsea Bauer, Lori Brinlee, John Chaney, Lindsey DeVries, Gay Deitrich-MacLean, Robert Gallen, Sandhya Nambiar, Alan King, MaryBeth McGavran, Kurt Michael, Sara Penning, Jonathan Rule, Douglas Scambler, Brandi Shaw, Katherine Stone, Maureen Sullivan and Ben Wilkowski. In addition, we are grateful to Richard Milich, Christine McKibbin, and several anonymous reviewers for their comments on previous versions of this manuscript.

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*, 519–533.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders, 4th edition, Text revision*. Washington, DC: Author.

Barkley, R. A. (2006). Attention-deficit/hyperactivity disorder: A handbook for diagnosis and treatment. In R. A. Barkley (Ed.), *Comorbid disorders, social and family adjustment, and subtyping* (3rd ed.). New York: Guilford.

Barkley, R. A., & Murphy, K. R. (2006). *Attention-deficit/hyperactivity disorder: A clinical workbook* (3rd ed.). New York: Guilford.

Burns, G. L., Boe, B., Walsh, J. A., Sommers-Flanagan, R., & Teegarden, L. A. (2001). A confirmatory factor analysis on the DSM-IV ADHD and ODD symptoms: What is the best model for the organization of these symptoms? *Journal of Abnormal Child Psychology, 29*, 339–349.

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

Hartung, C. M., Van Pelt, J. C., Armendariz, M. L., & Knight, L. A. (2006). Biases in ratings of disruptive behavior in children: Effects of sex and negative halos. *Journal of Attention Disorders, 9*, 620–630.

Jackson, D. A., & King, A. R. (2004). Gender differences in the effects of oppositional behavior on teacher ratings of ADHD symptoms. *Journal of Abnormal Child Psychology, 32*, 215–224.

McMahon, R. J., & Frick, P. J. (2005). Evidence-based assessment of conduct problems in children and adolescents. *Journal of Clinical Child and Adolescent Psychology, 34*, 477–505.

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*, 210–218.

Pelham, W. E., Fabiano, G. A., & Massetti, G. M. (2005). Evidence-based assessment of attention-deficit hyperactivity disorder in children and adolescents. *Journal of Clinical Child and Adolescent Psychology, 34*, 449–476.

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.

Sheskin, D. J. (2007). *Handbook of parametric and nonparametric statistical procedures* (4th ed.). Boca Raton, FL: Chapman & Hall/CRC.

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

Swanson, J. & Carlson, C. L. (1994). DSM-IV rating scale for ADHD and ODD. Unpublished manuscript.

Wolraich, M. L., Lambert, W., Doffing, M. A., Bickman, L., Simmons, T., & Worley, K. (2003). Psychometric properties of the Vanderbilt ADHD diagnostic parent rating scale in a referred population. *Journal of Pediatric Psychology, 28*, 559–568.