


Effects of Noninformational Color on the Reading Test Performance of Students with and without Attentional Deficits

Sydney S. Zentall, Janice A. Grskovic, James Javorsky, and Arlene M. Hall,1 Purdue University

Students with attentional deficits are better able to maintain attention to tasks if novelty is added, especially during later trials. In this study, we assessed generality to a standardized reading test when noninformational color was added to one of two alternate forms, counterbalanced for order of condition and form. Participants were 25 third- to fifth-grade students with and without attentional deficits. Students with attentional deficits: (a) read as accurately as their classmates with color added (during a first test administration and across sessions), (b) read worse in the black-white condition, and (c) improved reading accuracy during the second test administration with color added (i.e., all other groups showed a decline in performance).

Because the noninformational color did not change test difficulty (i.e., by altering the discriminability of task-relevant features), the implications of these findings may be more accurate assessment of the performance of students with attentional deficits.

One of the problems that students with attentional deficits bring both to the classroom and to the testing situation is a failure to maintain attention to lengthy tasks (i.e., there is a decline in performance over time, Zentall, 1985, 1986). Decreased performance over time indicates habituation to task novelty, especially when tasks involve practice or repeated information input (e.g., Shroyer & Zentall, 1986; Zentall, Falkenberg, & Smith, 1985). Especially when tasks are overly familiar, long, or difficult, students with attentional deficits tend to refocus their attention away from the task and toward what is salient or novel in their internal or external environments or generate stimulation through active responding.

We have theorized and reviewed evidence that students with attention deficit/hyperactivity disorder (AD/HD) need more stimulation or are more readily underaroused (Zentall, 1975; Zentall & Zentall, 1983). For example, evidence suggests that these students respond positively to increased levels of environmental stimulation (Zentall, 1980; Zentall & Zentall, 1976). It is also possible to help students with AD/HD maintain attention by placing novelty directly into their tasks, especially during later trials after habituation to the task stimuli has occurred. Color stim-
Color has been documented to produce better performance for students with AD/HD when it was added during later trials after spelling practice (Zentall, Zentall, & Booth, 1978) or after an initial exposure to traditional black lettering (Zentall & Dwyer, 1988; Zentall, 1989). In the latter study, the students with AD/HD outperformed students without attentional deficits when color was added late to those letters that were difficult to remember in spelling tasks (i.e., color plus information or relevant stimulation).

Because it is not always possible or practical to highlight relevant information (or justified in a test situation), we were interested in applications of color added late without highlighting information (color added only for stimulus value). To this purpose, we examined the effects of nonrelevant color on reading test performance of students with and without attentional deficits.

Students with disabilities are entitled to testing accommodations to compensate for poorer performance attributable to their disabilities (McDonnell, McLaughlin, & Morison, 1997). However, the use of accommodations that change the test standardization can limit the usefulness of the test scores. The identification of accommodations that maintain the test's validity is a critical need in the assessment field (Philips, 1995). In a review of the accommodation strategies used in large-scale group assessments by the National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Butler and Stevens (1997) identified the use of visual supports, such as color, as one of the prime forms of test accommodations. In the application of this visual support, it is prohibited to highlight or prompt the student’s attention to the relevant aspects of a test item (Indiana Department of Education, 1998). Thus, if it were possible to demonstrate improved performance on tests with no change in test difficulty, a more valid assessment of a student’s achievement gains could be obtained.

In summary, students with AD/HD have a theoretically based greater need for stimulation or habituate more rapidly to available stimulation. In support of theory, color stimulation has been demonstrated to improve the performance of these children, especially when it was added later into tasks to highlight relevant cues (e.g., to specific parts of letters that children typically fail to close in handwriting tasks, Zentall et al., 1985). What has not been demonstrated is whether color stimulation can be beneficial when it does not draw attention to relevant cues. We had demonstrated that nonrelevant color could be used to increase reading comprehension of students with AD/HD and reading disabilities, but only during the first third of 17 sessions (Belfiore, Grskovic, Murphy, & Zentall, 1996). These beneficial effects of color (embedded in a single-subject, within-series alternating treatments design) were not maintained when both the task and the color became less novel and were not assessed in comparison to students without disabilities.

Thus, the purpose of this study was to determine whether background color stimulation added to a reading test that did not require continued administrations would improve reading performance for students with attentional deficits more than for their classmates. The background stimulation selected was changes of color highlighting of whole lines of reading text. We predicted that highlighting would improve attention and thereby improve accuracy (i.e., by facilitating previously acquired reading responses of symbol to sound associations and sight vocabulary recognition); however, it would not be expected to facilitate the learning of new words or new understandings. These predictions are similar to those reported for psychostimulant medication. That is, Douglas, Barr, O’Neill, and Britton (1986) concluded that psychostimulant medication could not be expected to improve complex task performance, even though the primary deficits of failure to sustain attention and inhibit responses had been altered, unless remedial training was also provided for the secondary academic or cognitive deficits.

Method

Participants and Setting

Third-, fourth-, and fifth-grade students from five midwestern elementary schools were nominated by their teachers as either highly active and inattentive and not receiving special education services or as students from the same classrooms, who were similar in age and gender but average in activity and attention and not receiving special services. For the majority of schools, participating teachers nominated approximately three students in each group.

This procedure produced 34 students in 13 classrooms, all of whom consented to participate. Parents and teachers of the nominated students completed the Children’s Attention and Adjustment Survey - School Form (CAAS-S) and Home Form (CAAS-H) (Lambert, Hartsough, & Sandoval, 1990). This instrument is a 31-item rating scale with scales entitled Inattention, Impulsivity, Hyperactivity, Aggressiveness, Conduct Problems, AD/HD, and DSM-III-R AD/HD.
On the CAAS-School Form, the test-retest reliability of the DSM-III-R scale was \( r = .90 \), exceeding the minimum level for standardized diagnostic assessment instruments (see Aiken, 1985; Helmsdath, 1964; McDaniel, 1994; Nunnally, 1978; Salvia & Ysseldyke, 1995). The concurrent validity was established with the \textit{Pupil Behavior Rating Scale}, \( r = .51-.53 \) (Lambert & Hartsough, 1979) and the \textit{Connors Abbreviated Symptom Questionnaire}, \( r = .45-.66 \) (Connors, 1973). On the CAAS home form, the test-retest reliability of the CAAS-H DSM-III-R AD/HD scale, \( r = .75-.81 \), also met the minimal level (see Aiken, 1985; Helmsdath, 1964; McDaniel, 1994; Nunnally, 1978; Salvia & Ysseldyke, 1995). The concurrent validity was established with the \textit{Connors Abbreviated Symptom Questionnaire}, \( r = .65-.66 \), (Connors, 1973).

The DSM-III-R AD/HD scale, which was based on the \textit{Diagnostic and Statistical Manual of Mental Disorders - III-R} criteria (American Psychiatric Association, 1987), defined group placement. Even though the CAAS was normed using DSM-III-R AD/HD criteria, Lambert (1999) reported that the CAAS had diagnostic and research utility using DSM-IV AD/HD criteria (American Psychiatric Association, 1994). Participants were placed in the attentional deficits (AD) group if they had standard scores of 115 or more on the CAAS-S DSM-III-R AD/HD scale (1 SD above the mean was suggested as a minimum criterion score for AD/HD by Lambert et al., 1990, p. 23). Students scoring 104 or less were placed in the nondisordered comparison (NC) group. These procedures produced a sample of 25 students, 16 students with AD (3 girls, 13 boys) and 9 students in the comparison group (6 girls, 3 boys). The overall mean of the AD group was 127 (SD = 10.32), which represented the 96th percentile rank or almost 2 SD above the mean and differed from the overall mean of the NC group of 100.20 (SD = 6.38), which represented the 50th percentile rank, \( F(1, 23) = 70.59, p = .0001 \).

Confirmation of group status was established using home ratings by parents on the CAAS-H DSM-III-R AD/HD scale, which were obtained for 14 of the 15 participants in the AD group and 8 of the 9 participants in the NC group. On this parent scale, the AD group (\( M = 125.93, SD = 7.01 \)) also scored higher than the NC group (\( M = 97.00, SD = 6.48 \), \( F(1, 20) = 91.33, p = .0001 \). All students met criterion on both the parent and teacher scales, except for one participant, who received a score of 96 on the CAAS-S and a score of 128, or 2 SD above the mean, on the CAAS-H. We placed this participant in the AD group but ran our
data with and without this student with equivalent findings. Nine participants who were rated with scores from 105 to 114 were not included in the study. By eliminating these participants, we may have restricted the size of the NC group; at the same time, we eliminated scores that were marginal for group placement.

The two groups differed in gender composition, \( \chi^2 (1, N = 25) = 5.74, p = .017 \), and in ratings on the CAAS-Home Inattention scale, \( F(1, 20) = 79.23, p = .0001 \), and on the CAAS-School Inattention scale, \( F(1, 23) = 94.63, p = .0001 \), but not in grade or age (see Table 1). There were no differences between the groups on the \textit{Peabody Picture Vocabulary Test-Revised} (PPVT-R; Dunn & Dunn, 1981); however, a power analysis revealed that this test had a power of .34. By increasing the sample size to 44 participants per group, significant group differences would have been obtained on the PPVT-R. The PPVT-R assesses receptive language and is considered a measure of verbal aptitude unbiased by reading ability (Salvia & Ysseldyke, 1995).

Table 1. Demographic Equivalence Between Groups (Means and SD)

<table>
<thead>
<tr>
<th></th>
<th>AD</th>
<th>NC</th>
<th>( df )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>10.14</td>
<td>10.13</td>
<td>23</td>
</tr>
<tr>
<td>Grade</td>
<td>3.56</td>
<td>3.89</td>
<td>23</td>
</tr>
</tbody>
</table>
| CAAS-S|126.46  |100.20  |23       | 70.59***
| DSM-III-R AD/HD |       |        |   |
| CAAS-H|125.93  |97.00   |20       | 91.33***
| DSM-III-R AD/HD |       |        |   |
| CAAS-S|123.94  |94.33   |123      | 94.63***
| Inattention |       |        |   |
| CAAS-H|127.00  |97.38   |120      | 79.23***
| PPVT-R|104.00  |112.80  |23       | 2.71 |

Note. AD = attentional deficits; NC = normal controls; CAAS-S = School Form of DSM-III-R subscale of the Children's Attention and Adjustment Survey; CAAS-H = Home Form of DSM-III-R subscale of the Children's Attention and Adjustment Survey; PPVT-R = Peabody Picture Vocabulary Test - Revised.

***p < .0001.
Measures

The Gray Oral Reading Test 3 (GORT-3; Wiederholt & Bryant, 1992) is an individually administered standardized test of reading with alternate forms (A and B), which measures oral reading speed (rate), decoding (accuracy), and comprehension. The GORT-3 is sensitive to treatment effects and useful for research, due to the availability of equivalent forms (Wiederholt & Bryant, 1992). The GORT-3 consists of 13 reading passages of increasing difficulty. Test performance is measured by the Oral Reading (total test score) and by Accuracy, Rate, and Comprehension subscales. Subscale scores are reported in standard scores, M = 10 and SD = 3. Using the GORT-3 scoring procedures, a 5-point system converted the time to read the story to a Rate subscale score and the number of oral reading errors (deviations from print for errors of addition, repetition, self-correction, omission, transposition, and pauses of 5 seconds or more) to an Accuracy subscale score. The Comprehension subscale score was derived from answers to multiple-choice questions, which were read aloud by the examiner with the child following along (i.e., reducing impulsive or inattentive responses). The comprehension questions were designed to measure not only literal comprehension and recall of facts but also inferential, critical, and affective reading skills. For a test like the GORT-3 to be considered reliable, its reliability coefficients must approximate or exceed .80 in magnitude, with coefficients of .90 or above considered desirable (see Aiken, 1985; Helmstader, 1964; McDaniel, 1994; Nunnally, 1978; Salvia & Ysseldyke, 1995). Reliability coefficients of the GORT-3 fall within the desirable range (test stability median coefficient, r = .90; alternative form reliabilities for the Rate, r = .82, Accuracy, r = .80, and Comprehension, r = .62, subscales).

We received permission from the publisher to reproduce passages from the GORT-3 on card stock (22 by 28 cm), as they typically appear in the GORT-3 Student Book, and to spiral-bind each of the two forms. It was administered individually by one of three doctoral students in special education, who were naive to the participants’ group status. Test administrators received training in GORT-3 standardized administration procedures and then completed a written exam, which had been developed from the procedural steps presented in the test manual. Additionally, the procedural integrity of test administration was assessed by the trainer through the use of a procedural checklist. Training and assessment continued until a mastery criterion of 95% was reached on both measures.

Independent Variable

Two levels of the novelty condition (black-on-white and color-added-late) were constructed. The black-on-white (BW) condition consisted of reading passages of black lettering on white card stock, exactly as they appear in the GORT-3 Student Book. The noninformational color, added late (CL) condition appeared the same as the BW condition for the first third of the text of each story. But in the middle third of each story, the lines of type were highlighted with pastel colors (either sky blue, gray, apricot, orchid, pink, spring green, or sea green) by drawing over complete sentences with magic markers, and in the last third of each story, lines of bold colored highlighting (either fiery orange, bittersweet, shocking pink, goldenrod, electric blue, marigold, or hot magenta) were drawn over the sentences. Thus, depending on the story length (difficulty), noncolored black lettering may have been from one to five lines, followed by one to five lines of pastel highlighted black lettering, and one to five lines of bold highlighted lines. Color changes or increments in the text (white to pastel color to bold color) were made contingent upon the end of sentences that fell within the line that was closest to the end of a third of a passage. These increments were made to offset the expected increased habituation to task stimuli over time, especially for children with attentional disorders.

Experimental Setting and Procedures

Two testing sessions, each 25-30 minutes, took place in each child’s school during the school day and were separated by two weeks time. Testing rooms varied across schools (media center, speech therapy room, or a corner of a gymnasium) with space varying from about 1.2 m by 1.8 m to approximately 6 m by 6 m. At each site, an experimenter tested a student one-on-one at a table.

Condition (black-on-white and color-added-late) and equivalent test form order (A and B) were counterbalanced with the constraints that approximately half the number of students in each group were randomly assigned to receive the color first (half of these with form A and half with form B) whereas the other half received the black-on-white test first, with a similar counterbalancing of form order. During each initial session, the first condition and form of the GORT-3 were administered, followed by the PPVT-R. During the second session, the alternate condition and form of the GORT-3 were presented. Both tests were administered according to the standardized procedures presented in their man-
uals. On the GORT-3, students began reading with the story that corresponded to their grade placement and stopped reading when they had reached the ceiling. At the end of each testing session, participants received a token of appreciation (e.g., sports card, sticker).

Results

Mixed-design (Tabachnick & Fidell, 1996) analyses of variance (ANOVA, GLM model adjusts for unbalanced groups) were used to examine the independent factors of group (AD, NC) and condition order (CL first, BW first), as well as the repeated factor of condition (CL, BW) for Rate, Accuracy, and Comprehension subscale scores. These scores were used, rather than the Oral Reading quotient, because they were independent scores (i.e., not composites). Prior to this analysis, we assessed the assumptions of the analysis of variance. That is, kurtosis and skewness for all variables were assessed within the normal limits for the total sample. The normality plots of the dependent measures, as well as the covariates, yielded a normal distribution for all variables, as well as, no within-cell outliers at \( p = .001 \) (Anderson & McLean, 1974). In testing homogeneity of variance between the groups, a \( F_{\text{max}} \) test was used. This test divides the largest standard deviation of the AD group (AD: Comprehension; \( SD = 4.16 \)) by the smallest standard deviation of the NC group (NC: Rate; \( SD = 1.90 \)) to compute a \( F_{\text{max}} = 2.19 \). In order to conclude that there were significant differences between the groups, the calculated \( F_{\text{max}} \) would have to be greater than 8.89; thus, we can conclude that the variances between the two groups were homogeneous. We also assessed the cells of the design and found demographic equivalence in the nonsignificant group by condition-order interactions for age, \( F(1,21) = .98, p = .332 \), hyperactivity ratings, \( F(1, 21) = .24, p = .631 \), and PPVT-R scores, \( F(1, 21) = 4.31, p = .051 \). We documented design equivalence in a Group x Gender x Condition Order interaction, but subsequently dropped gender from the model, because main effects and interactions were not found for our dependent variables (Accuracy: \( F(4, 16) = .18, p = .736 \); Comprehension: \( F(4, 16) = .69, p = .604 \); Rate: \( F(4, 16) = .89, p = .487 \). Using the main mixed-design model, group differences were found for Rate, Accuracy, and Comprehension subscale scores. These effects indicated that students with AD read less accurately (\( M = 9.81, SD = 3.67 \)) than comparisons (\( M = 13.11, SD = 2.67 \)); read at a slower rate (\( M = 9.88, SD = 3.63 \)) than comparisons (\( M = 13.11, SD = 1.90 \)); and had lower comprehension (\( M = 9.00, SD = 4.16 \)) than comparisons (\( M = 12.00, SD = 2.06 \)). The statistical power for each of these analyses was adequate: (a) Accuracy subscale, statistical power = .69; (b) Rate subscale, statistical power = .80; and (c) Comprehension subscale, statistical power = .63.

We statistically partialed out the effects of verbal aptitude and reran each of the three analyses for the following reasons: Homogeneity of regression slopes of the PPVT-R covariate for Rate, Accuracy, and Comprehension subscale scores were found to be nonsignificant in the group by PPVT-R interaction—supporting the appropriateness of this covariate. With the analysis of covariance (ANCOVA) design, each main effect of group was lost (see Table 2), with PPVT-R scores accounting for significant amounts of the variance in each model of Rate, Accuracy, and Comprehension. For this reason, we report the ANCOVA and the least square means that have been adjusted by the PPVT-R scores.

In the overall analysis a three-way interaction (Condition x Group x Order) was obtained for Accuracy (see Figure 1 and Table 2). This interaction resulted from the fact that only one of the four groups (AD: BW first/CL second) showed an improvement in reading accuracy during the second administration, as determined by paired comparison t-tests of differences, diff = +1.5, \( p = .026 \). The other groups showed a decline in

Figure 1. Mean accuracy of oral reading for children with attentional deficits (AD) and for nondisordered comparisons (NC) by condition order: color first and black-white second (CL 1st) and black-white first and color second (BW 1st). Scores are on a standard scale with a \( M \) of 10 and a SD of 3.
Table 2. Analyses of Covariance of Condition Effects on the Reading Test Performance of Students with and without Attention Deficits

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Accuracy</th>
<th>Rate</th>
<th>Comprehension</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Between subjects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPVT-R covariate</td>
<td>1</td>
<td>6.29*</td>
<td>7.13*</td>
<td>14.50**</td>
</tr>
<tr>
<td>Group (G)</td>
<td>1</td>
<td>2.08</td>
<td>2.47</td>
<td>1.97</td>
</tr>
<tr>
<td>Order (O)</td>
<td>1</td>
<td>0.98</td>
<td>1.23</td>
<td>2.71</td>
</tr>
<tr>
<td>G x O</td>
<td>1</td>
<td>0.51</td>
<td>0.01</td>
<td>0.60</td>
</tr>
<tr>
<td>Error</td>
<td>19</td>
<td>(283.55)</td>
<td>(302.78)</td>
<td>(177.84)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Within subjects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition (C)</td>
<td>1</td>
<td>0.25</td>
<td>0.02</td>
<td>1.02</td>
</tr>
<tr>
<td>C x G</td>
<td>1</td>
<td>6.78*</td>
<td>0.99</td>
<td>0.13</td>
</tr>
<tr>
<td>C x O</td>
<td>1</td>
<td>0.14</td>
<td>0.11</td>
<td>0.60</td>
</tr>
<tr>
<td>C x G x O</td>
<td>1</td>
<td>11.56**</td>
<td>0.89</td>
<td>0.61</td>
</tr>
</tbody>
</table>

Note. Values enclosed in parentheses represent mean square errors.
* p < .05; ** p < .01.

performance during the second administration (AD: CL first/BW second, diff = -0.38, p = .481); this decrease was significant for one of the comparison groups (NC: CL first/BW second, diff = -0.8, p = .016, and NC: BW first/CL second, diff = -1.0, p = .092).

In an examination of group differences in just the first administration of the GORT-3 (see Figure 1, First Administration), which represents the performance that would have been assessed in a typical testing situation, significant differences were found among the four groups, F (3, 21) = 3.39, p = .037. A follow-up Tukey's Test demonstrated that students with AD scored worse (p < .05) than their classmates only in the traditional black-and-white version of the GORT-3. In other words, the students with AD demonstrated equivalent performance to that of the comparison group in the CL condition.

A lower level two-way group by condition interaction for Accuracy was also yielded in the ANCOVA data, F (1, 16) = 7.79, p = .013. The mean adjusted scores of students with AD indicated that they performed with greater accuracy in the CL condition (M = 10.94, SEM = 0.73) than in the BW condition (M = 9.91, SEM = 0.75)—relative to the accuracy of comparison students, who scored better in the BW condition (M = 12.25, SEM = 1.01) than CL condition (M = 11.98, SEM = 1.01). Tukey's analyses indicated no group differences in reading accuracy in the CL condition; however, a group difference (p < .05) in the BW condition was observed. These findings indicate that the CL condition may have normalized the accuracy performance of participants with attentional deficits. In other words, the students with AD demonstrated equivalent performance to that of their classmates only in the CL condition.

To assess the significance of these results, we derived effect size scores (Mcl - Mbw) / SDbw = Effect Size (Glass & Hopkins, 1984, p. 236) using the mean of the Accuracy subscale score. The standard deviations of the Accuracy subscale scores were not significantly different between AD and NC groups. The gain from the color-added condition over the traditional BW condition for the AD group was estimated to be .31 z-score standard deviations, indicating an improvement of 12 percentile ranks. These effects are modest, not dramatic, but clearly present. That is, the traditional black text on white background reading test would place students with AD at the 50th percentile rank, and an effect size of .31 would move those children to the 62nd percentile rank (Glass & Hopkins, 1984, p. 64). An increase of 12 percentile ranks would be equivalent to a modest improvement from 100 to 105 standard score points. The difference between the CL and BW conditions for the control group was estimated to be 0 standard deviations, which indicated a negligible treatment effect for the control group.

Also in the analysis of the Accuracy subscale scores of the AD group, we examined whether the participants' scores changed from one performance range to another (see Table 3). Ranges were defined as: (a) below average ≤ 7 subscale score, (b) average = 8–12 subscale scores, and (c) above average ≥ 13 subscale score. We found that 78% of the NC participants remained in the same performance range, while 11% increased and 11% decreased in their performance ranges. At the same time, 50% of the AD participants remained in the same range, while 40% increased their performance range and 10% decreased.

Whereas color improved the reading accuracy of students with AD (relative to their own performance in traditional black-and-white and to the performance of comparison students), it did not alter their reading rate nor their comprehension of the material, as indicated in the ANCOVA data (see Table 2).
Table 3. Accuracy Scores of Participants for CL and BW Conditions

<table>
<thead>
<tr>
<th>Student Group</th>
<th>BW</th>
<th>SS</th>
<th>Percentile</th>
<th>Rating</th>
<th>CL</th>
<th>SS</th>
<th>Percentile</th>
<th>Rating</th>
<th>SS Diff</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AD 6</td>
<td>50</td>
<td>average</td>
<td>8 25</td>
<td>average</td>
<td>+.66</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AD 10</td>
<td>50</td>
<td>above average</td>
<td>14 91</td>
<td>above average</td>
<td>+.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AD 13</td>
<td>84</td>
<td>above average</td>
<td>12 75</td>
<td>average</td>
<td>-.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AD 11</td>
<td>63</td>
<td>average</td>
<td>12 75</td>
<td>average</td>
<td>+.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AD 3</td>
<td>1</td>
<td>very poor</td>
<td>2 1</td>
<td>very poor</td>
<td>-.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AD 12</td>
<td>75</td>
<td>average</td>
<td>13 84</td>
<td>above average</td>
<td>+.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AD 12</td>
<td>75</td>
<td>average</td>
<td>11 63</td>
<td>average</td>
<td>-.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AD 8</td>
<td>25</td>
<td>average</td>
<td>9 37</td>
<td>average</td>
<td>+.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AD 15</td>
<td>95</td>
<td>superior</td>
<td>15 95</td>
<td>superior</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AD 5</td>
<td>5</td>
<td>poor</td>
<td>7 16</td>
<td>below average</td>
<td>+.66</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AD 10</td>
<td>50</td>
<td>average</td>
<td>9 37</td>
<td>average</td>
<td>-.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AD 9</td>
<td>37</td>
<td>average</td>
<td>8 25</td>
<td>average</td>
<td>-.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AD 5</td>
<td>5</td>
<td>poor</td>
<td>7 16</td>
<td>below average</td>
<td>+.66</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AD 12</td>
<td>75</td>
<td>average</td>
<td>13 84</td>
<td>above average</td>
<td>+.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AD 11</td>
<td>63</td>
<td>average</td>
<td>14 91</td>
<td>above average</td>
<td>+1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC 11</td>
<td>63</td>
<td>average</td>
<td>11 63</td>
<td>average</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC 12</td>
<td>75</td>
<td>average</td>
<td>11 63</td>
<td>average</td>
<td>-.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC 16</td>
<td>98</td>
<td>superior</td>
<td>14 91</td>
<td>above average</td>
<td>-.66</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC 11</td>
<td>63</td>
<td>average</td>
<td>11 63</td>
<td>average</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC 15</td>
<td>95</td>
<td>superior</td>
<td>15 95</td>
<td>superior</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC 12</td>
<td>75</td>
<td>average</td>
<td>13 84</td>
<td>above average</td>
<td>+.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC 9</td>
<td>37</td>
<td>average</td>
<td>10 50</td>
<td>average</td>
<td>+.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC 18</td>
<td>99</td>
<td>very superior</td>
<td>17 99</td>
<td>very superior</td>
<td>-.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC 10</td>
<td>50</td>
<td>average</td>
<td>11 63</td>
<td>average</td>
<td>+.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+.27</td>
</tr>
</tbody>
</table>

Note. AD = attention deficits; NC = normal controls; SS = Scaled Scores (M = 10, SD = 3), Diff

Discussion

This study assessed the performance effects of adding noninformational color highlighting to the last two thirds of text in a reading test for students with attentional deficits (AD). We anticipated more accurate assessment of reading skill for students with attentional deficits than has been demonstrated on traditional reading tests with black text presented on a white background.

In this study, we found that students with AD read slower, less accurately, and had lower comprehension than comparisons. However, group differences for comprehension and reading rate did not remain when we controlled for the effects of verbal IQ (PPVT-R). Reading accuracy was the variable unaffected by differences in verbal IQ and consistently sensitive to treatment. For example, during the first test administration, we found overall that students with AD read as accurately as their classmates only in the condition with color added late. During the second test administration, we also found an improvement in reading accuracy for that group of students with AD whose second administration was presented in the color condition (i.e., the three other groups showed a decline in performance on the second administration). This decline in accuracy for most of the children (except for the AD group with color added) indicates that repeated exposure to a reading decoding task required sustained attention.

Across both first and second sessions, the simple effects analyses of the group-by-condition interaction indicated that students with AD read as accurately as their classmates in the color-added late condition, and their accuracy was lower than comparisons in the BW condition (i.e., their performance was normalized by the color added). We concluded that these effects were clinically important by increasing students' Accuracy scores an average of 12 percentile ranks and practically important by improving the performance range of 40% of the AD group versus 11% of the NC group, with comparable percentages of decreased performance (10% vs. 11%).

Overall, our findings indicated that noninformational color added late can improve reading accuracy. Although some of our prior work produced even stronger findings by highlighting relevant information, it is important, as shown in the present study, that added color can have significant effects, even when it does not involve highlighting information. These findings do not explain, however, why the beneficial effects of added color were demonstrated for Accuracy and not for Comprehension. We did expect improved recall of simple information. However, the GORT-3 comprehension questions required inferential thinking (e.g., why, how, what next) in addition to simple recall of information. Inferential thinking may be less affected by attentional manipulations than recall of information. Our conclusion that recalled information may be more sensitive to manipulations of added color is supported by findings from a single-subject study assessing comprehension with literal recall of short-answer questions (e.g., who, what, when, where) (Belfiore et al., 1996). In that study, students with AD/HD read passages

COLOR ADDED TO READING
silently and demonstrated a 43% increase in comprehension during the first third of 17 sessions in the color condition. Unfortunately in this study, it was not possible to reanalyze only the recall questions, due to differences in the forms used, number of questions per story, and lack of normative information at this level of analysis.

It is also possible that the lack of effects for Comprehension in the present study could be attributed to the lower reliability of the Comprehension subscale and to the fact that the GORT-3 required students to divide their attention between speed of oral decoding and comprehension. The test administrator asked students to read the stories aloud “as quickly” and “as well” as they could and when finished to answer some questions about the reading material. That is, the instructions focused primarily on the decoding task and only secondarily on comprehension.

The findings of our study are limited by the fact that the study assessed a relatively small school-based sample, and generality to a clinical sample of children with a diagnosis of AD/HD therefore is unknown. Replication with a clinical sample may be a direction for future research, although prior research has documented equivalent effects from color stimulation for clinical samples of children as for children within general education settings. It might also be argued that a design involving a small sample size does not warrant a ready interpretation of significant effects. However, Rosenthal and Gaito (1963) have indicated that given two experiments with the same level of significance, the experiment with the smaller sample size would have produced more impressive results.

The strengths of this study are related to our findings that noninformational color added late to the background of reading text can improve reading accuracy only for students with attentional deficits. Furthermore, there is some indication from prior work that it can improve comprehension for students with AD/HD, when comprehension is defined as recall of information. The effect of nonrelevant color was especially beneficial (a) when it was added later after an initial session or exposure to traditional black lettering on white and (b) on initial administrations of color before habituation to the color occurred.

Implications

The diagnostic and assessment implications of these findings are broadly related to an accurate, practical, and unbiased assessment of the reading abilities of students with attentional deficits. That is, some read-

COLOR ADDED TO READING

ing assessments may not represent these students “true” abilities or knowledge, because true performance is confounded by the attentional demands of the test. This modification is especially appropriate for curriculum-based assessment (CBA), which involves an informal measurement of expected curricula outcomes (Taylor, 1997). To compensate for the effects of attentional disorders, teachers can make modifications for testing situations to attain a truer measure of reading abilities. These modifications are appropriate for curriculum-based assessment, which is designed by teachers to assess individual student strengths and weaknesses (Fuchs & Fuchs, 1988; Taylor, 1997). An assessment of actual reading performance under accommodated conditions (such as the color added late test) is useful for determining a student’s instructional level within a language arts curriculum and for assessing the efficacy of reading interventions and instructional programs (year-end gains). This practical and effective accommodation can easily be used in educational contexts even by students who could learn to highlight their own reading materials in ways that help guide and sustain their attention. Note that these attentional manipulations will not teach children how to decode or address problems attributable to dyslexia (Lyon, 1992), but can improve the accuracy of already learned skills. In other words, manipulations of color address the performance deficits of children with attentional problems and not their skill deficits.

However, we also need to consider a second type of performance—reading performance under unaccommodated classroom conditions (typical performance). Performance under typical conditions is useful in predicting general classroom functioning without accommodations and determining eligibility for services (Zentall & Javorsky, 1995). In recognition of the importance of assessing true performance separately from typical performance, special educators and diagnosticians (a) administer tests while the student is on and off medication, (b) administer the test in shorter time periods, (c) provide external contingencies, or (d) alter the administration procedure. Some of these alternatives may not be medically appropriate (e.g., placing the child on/off medication for assessment purposes). Other alternatives may be practical and even provide evidence for possible accommodations in the classroom (e.g., extending time, using incentives) but still violate the standardization of the test and the applicability of the test’s norms. For these reasons, additional research is needed to determine the predictive validity of noninformational color added late for assessing true performance.
Given the widespread efforts directed toward assessing the efficacy of psychostimulant medication, this study uniquely examines variables related to school success for students with attentional deficits. Our results in this area are important practically with respect to the types of test and assessment procedures that yield optimal reading performance without altering task-relevant features. The findings are also theoretically important by contributing to the literature documenting differential performance gains from visual stimulation added to the tasks of students with attentional disorders, hyperactivity, and AD/HD. Past research has documented that stimulation can be added to relevant cues to direct students' attention to information and therefore to increase learning. In the present study, color stimulation was added without information to increase the performance of already learned skills, similar to the more general attention enhancing effects of stimulant medication.

References

COLOR ADDED TO READING
Phillips, S. E. (1995, November). All students, same test, same standards: What the new Title I legislation will mean for the educational assessment of special education students. Oak Brook, IL: North Central Regional Educational Laboratory.
Parent, Teacher, and Self-Report of Problem and Adaptive Behaviors in Children and Adolescents with Asperger Syndrome

Gena P. Barnhill, Taku Hagiwara, Brenda Smith Myles, Richard L. Simpson, Megan L. Brick, and Deborah F. Griswold, University of Kansas

The present study examined perceptions of the social problems and adaptive behaviors of children and youth with Asperger Syndrome. Parents and teachers used the Behavior Assessment System for Children (BASC) (Reynolds & Kamphaus, 1992) to evaluate 20 children and youth with Asperger Syndrome. In addition, the 20 students using the BASC self-report instrument evaluated their social problems and adaptive behavior. Findings are discussed relative to better understanding and planning for the needs of children and youth with Asperger Syndrome and their families.

In his 1944 seminal paper, Asperger (1991) described five children who developed language at an appropriate age, but who demonstrated odd behaviors, were resistant to change, used pedantic and stereotyped speech and displayed deficient social behavior. Asperger argued that while the children exhibited “typical autistic behaviors,” they were worthy of a diagnostic category of their own that he called “autistic psychopathy” (later termed Asperger’s Syndrome by Wing, 1981, p. 115). Although Asperger Syndrome was initially described over five decades ago, the condition was virtually unnoticed until the 1980s when Wing (1981) published a comprehensive clinical account of 34 case histories of individuals he diagnosed with Asperger Syndrome, ranging in age from 5 to 35 years. Wing (1981) described seven features of individuals with Asperger Syndrome: lack of empathy; naivété; inappropriate, one-sided interactions; little or no ability to form friendships; pedantic, repetitive speech; poor nonverbal communication; intense absorption in certain obscure subjects; and clumsy, stereotyped motor movements. Frith (1991) and Gillberg (1989), among other have subsequently confirmed these features.

The most widely used diagnostic criteria for Asperger Syndrome (qualitative impairment in social interaction; circumscribed interest; restricted, repetitive, and stereotyped behaviors, interests, and activity; and no clinically significant general delay in language) have been included in the Diagnostic and Statistical Manual of Mental Disorders-Fourth Edition (DSM-IV) (American Psychiatric Association, 1994) and th