Vestibular and Visual Rotational Stimulation as Treatment for Attention Deficit and Hyperactivity

(occupational therapy, optokinetic stimulation, semicircular canal, sensory stimulation)

L. Eugene Arnold, David L. Clark, Larry A. Sachs, Stephanie Jakim, Chris Smithies

Previously published studies report an improvement in hyperactivity following exposure to vestibular semicircular canal stimulation under eyes-open, lights-on conditions (conditions that provide visual feedback). To separate the effectiveness of vestibular stimulation from that of visual stimulation, 30 primary school children who met the criteria for having attention deficit disorder with hyperactivity were enrolled in a split-sample Latin square crossover study. The effects of just vestibular stimulation and of just visual stimulation were compared with the effect of combined vestibular and visual stimulation. Behavior ratings showed significant improvement at the end of the last treatment and at follow-up one year later; this is not easily explained by statistical regression, history, or the placebo effect. The most improvement was with solitary vestibular stimulation, which showed large effect sizes; however, differences from the other two conditions failed to reach significance at traditional p levels.

A previous study showed significant improvement on teachers' ratings of 18 hyperkinetic children following systematic eyes-open, lights-on, rotatory stimulation to each of the three pairs of semicircular canals; this improvement is significant compared with a contact control condition (1). The rotatory semicircular canal stimulation was provided by a swivel chair with head cocked in the positions described below. Stimulation was provided twice a week for four weeks in a crossover design with an equal number of control contact sessions. The treatment advantage was restricted to the ten children under 10 years old without undersocialized aggressive features. Because that study used eyes-open rotation in a lighted room (thus providing visual feedback), the possibility that visual rotational (optokinetic) stimulation could account for a significant part of the benefit cannot be ruled out. The following is a study designed to find the relative importance of the vestibular component separate from that of the visual component of sensory stimulation.

Method

Children who were selected from kindergarten through third grade of five public schools in three school districts by teacher rating were studied in a split Latin square crossover design (see Table 1). Children were assigned to a three-condition series that included either a) control condition, b) combined visual and vestibular stimulation, and c) vestibular stimulation alone or a) control condition, b) combined visual and vestibular, and c) visual stimulation alone. Each series lasted 12 weeks: 4 weeks on each condition, two sessions per week.

The sample was nonclinical and was selected with the cooperation of the five elementary schools. Each child in the sample had to have the following criteria:

- Diagnostic and Statistical Manual—III (DSM-III) draft criteria for attention deficit disorder (ADD) with hyperactivity HA (at the time the project was initiated, only the semifinal DSM-III draft was available, which required two impulsiveness items and the gen-

L. Eugene Arnold, MEd, MD, is a professor, Pediatrics and Psychiatry Departments, David L. Clark, PhD, is an associate professor, Anatomy Department, and Larry A. Sachs, PhD, is a senior research associate; all at Ohio State University, Columbus, OH 43210.
Table 1  
Split-Sample Latin Square Design

<table>
<thead>
<tr>
<th>Order</th>
<th>1st 4 wks</th>
<th>2nd 4 wks</th>
<th>3rd 4 wks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TAV</td>
<td>B</td>
<td>Vestib</td>
</tr>
<tr>
<td>2</td>
<td>TAV</td>
<td>B</td>
<td>Vis</td>
</tr>
<tr>
<td>3</td>
<td>B</td>
<td>Vestib</td>
<td>TAV</td>
</tr>
<tr>
<td>4</td>
<td>B</td>
<td>Vis</td>
<td>TAV</td>
</tr>
<tr>
<td>5</td>
<td>Vestib</td>
<td>TAV</td>
<td>B</td>
</tr>
<tr>
<td>6</td>
<td>Vis</td>
<td>TAV</td>
<td>B</td>
</tr>
<tr>
<td>7</td>
<td>TAV</td>
<td>Vestib</td>
<td>B</td>
</tr>
<tr>
<td>8</td>
<td>TAV</td>
<td>Vis</td>
<td>B</td>
</tr>
<tr>
<td>9</td>
<td>B</td>
<td>TAV</td>
<td>Vestib</td>
</tr>
<tr>
<td>10</td>
<td>B</td>
<td>TAV</td>
<td>Vis</td>
</tr>
<tr>
<td>11</td>
<td>Vestib</td>
<td>B</td>
<td>TAV</td>
</tr>
</tbody>
</table>

- TAV, tactile-auditory-visual stimulation (control); B, both visual and vestibular stimulation; Vestib, vestibular stimulation alone; Vis, visual stimulation alone. 
- The whole 12-order matrix was repeated twice more to accumulate the total sample of 36 subjects. 
- Actual assignment to order of conditions within each series was random and not necessarily in the numbered sequence shown.

Several symptoms of inattentiveness and hyperactivity:
- a rating of 24 or more on the first six items of the Davids Hyperkinetic Rating Scale (2);
- a mean item rating of 1.5 or more for the hyperkinetic factor of the 39-item Conners Teacher Rating Scale (3), using the hyperkinetic factor derived from a previous nonclinical sample of 225 unselected first graders (4);
- an age of 9 years or below;
- normal intelligence as evidenced by either school records of group testing or individual testing or by a Draw-A-Person mental age or other assessment done for this purpose (the school psychologist assisted with this screening); and
- written parental consent and verbal child assent after the project was explained.

The mean and median age of the sample was 7 years, with a range of 5 to 9 years. There were 27 boys and 10 girls who completed the study. Whenever possible, the subjects were watched for school and severity of baseline behavioral ratings and then randomly assigned to either the vestibular-only or the visual-only series. Each series had the combined vestibular and visual rotatory stimulation with eyes open and lights-on (similar to the treatment in the original study) as one of the conditions.

Twice a week, at each of eight consecutive sessions (over a four-week period), the subject, wearing goggles with a clear lens, was seated in a rotary chair with lap belt restraint and instructed to tilt his or her head forward about 30°, thus placing the horizontal semicircular canals in the horizontal plane. The chair was accelerated clockwise manually to 150°/sec (25 rpm), held at that angular velocity for one minute, and then impulsively stopped in less than one second. The subject's head was held in the same position as the second spin, which was the original study was considered unsatisfactory because it was not very convincing: the child either rotated the chair while standing be-
side it or sat in the stationary chair and played games with the investigator. The validity of the control condition was supported by the fact that parents' ratings showed a placebo effect, whereas teachers' ratings did not; this pattern is similar to the results obtained in well-controlled drug studies with matched placebos.

Nevertheless, we had a great desire to find a more convincing sham condition. Therefore, we contrived the following structured control session: a) subject has air “poofed” onto the back of his or her hand for one minute, using a large rubber bellows (tactile stimulation), b) subject sits in the stationary chair for four minutes, keeping his or her eyes closed, wearing opaque goggles, and listening to the optokinetic drum spin around (auditory stimulation), and c) subject looks through a stereoscopic slide projector for two minutes (visual stimulation). It was presumed that none of these sensory stimuli alone would have a significant effect on the child's subsequent behavior. This condition was called TAV (tactile-auditory-visual stimulation). During the explanation prior to the study, the subjects and their parents merely were told that different aspects of the original stimulation were being tried separately to see which of them was the most effective element; these aspects included the sensation of spinning, the sight of spinning, the sound of the spinning, the visual training required by eye fixation, and the feel of the air breezing over the skin that spinning produces.

To facilitate sample retention, the rotating chair and surrounding optokinetic drum were built into a van, which was driven to the subjects' schools. This overcame parental objections about inconvenience or the stigma of bringing the nonpatient child to the psychiatry clinic twice a week.

The following measures were collected at screening (baseline 1), prior to any treatment (baseline 2), every 2 weeks for 12 weeks of treatment, and at the one-year follow-up:

- Davids Hyperkinetic Rating Scale by teachers;
- Conners 39-item teacher rating scale;
- a checklist of behavioral criteria for ADD with HA taken from the DSM-III provisional draft, filled out by teachers;
- Davids Hyperkinetic Rating Scale filled out by parents; and
- Conners short form Parents' Questionnaire.

A Bender-Gestalt test was performed at baseline and at the end of each of the one-month treatment periods.

When each rating was due, reminders were sent to teachers and parents. However, it soon became apparent that reminders sent to disinterested parents caused the parents to withdraw their children from the study. The selective attrition seemed threatening enough to sample representativeness that a revision in strategy seemed in order. Because the teachers' ratings were considered the more valid and important measure of behavioral effect, we decided that sample retention with complete teacher ratings was preferred over a decimated sample with both teachers' and parents' ratings. Accordingly, when we stopped reminding parents, the dropout rate ceased (except for one subject whose family moved to another city). Because data from so many parents were missing, only teachers ratings were analyzed and reported.

To check for statistical regression to the mean (considered necessary because the subjects were selected partly on the basis of scores on the rating scales that were used as dependent variables), we ran correlations between baseline 1 score and the amount of change from baseline 1 to baseline 2 scores. Significant correlations would suggest a significant amount of statistical regression. As a partial check on the effect of history, we made correlations between the length of time from baseline 1 to baseline 2 and the amount of change. The length of time between baseline 1 (screening) and baseline 2 (pretreatment) varied from two to over four weeks: a significant correlation between amount of change and time elapsed would suggest a significant history effect (maturation?).

The primary statistical procedure for assessing relative treatment effects was that of a balanced Latin square design (see Table 1). This procedure extracts and excludes sequence and carryover effects, thus testing treatment effects independently of these two factors. Separate sets of Latin square analyses were performed for the vestibular-alone series and for the visual-alone series for each of the dependent variables. A two-way analysis of variance with repeated measures (vestibular-alone series/visual-alone series) was done on the overall long-term change from baseline 2 to follow-up. For each series, paired t tests compared the behavioral scores at baseline 2 to the scores at the end of all 12 weeks of treatment.

Results

Of the 42 children who started the study, 37 completed all three treatment conditions. Unfortu-
nately, the dropouts affected the Latin square in such a way that only 30 subjects could be matched for the balancing order inherent in the Latin square design: 12 subjects in the vestibular-alone series and 18 subjects in the visual-alone series. Therefore, statistical comparisons of the three treatments conditions in the crossover were made only on these 30 subjects, but up to 37 were used for other statistical analyses. All \( p \) values reported are two tailed.

Correlations between \textit{baseline} 1 score and amount of change from \textit{baseline} 1 to \textit{baseline} 2 scores were not significant. Neither were any of the correlations between amount of change from \textit{baseline} 1 to \textit{baseline} 2 scores and length of time elapsed.

Mean behavioral ratings at the various checkpoints are displayed in Figures 1–3 for the vestibular-alone series and in Figures 4–6 for the visual-alone series. The mean changes under each condition are displayed in Table 2. A higher score on the rating scale is more symptomatic. The numbers displayed in Table 2 are the mean differences obtained by subtracting the rating scale score at the end of the condition from the score at the beginning of that condition. Therefore, a positively signed difference indicates improvement. Bender-Gestalt Koppitz scores did not correlate with behavior ratings \((r = .03\) and \(.04\) for the two scales), although comparison of vestibular-alone Bender scores to those on the other two conditions shows a large effect size (see Table 2).

The two-way analysis of variance using pooled subjects showed that the improvement from \textit{baseline} 2 to follow-up was significant at \( p < .005 \) for the Davids Hyperkinetic Rating Scale and at \( p < .05 \) for the
The sum of changes from baseline 2 (2nd baseline) to the end of all 12 wks of treatment was nonsignificant. Lower score is better; the mean ratings shown for the fourth week of each 4-week condition in the Latin square crossover were obtained by subtracting from 2nd baseline the mean adjusted change during that condition (shown and explained in Figure 2).

Latin square analysis of the relative improvement during each of the three conditions (control, combined vestibular and visual stimulation, and solitary visual or vestibular stimulation) showed a different pattern by series (see Table 2). In the vestibular-alone series, the vestibular-alone condition showed a trend of advantage that was not matched by the visual-alone condition in its series. Improvement was greater following vestibular-alone condition than following combined or TAV conditions on all three scales (see Figures 1-3). The trend is quantified best by the effect size (explained below), but even by traditional probability tests improvement approaches significance on the Davids Hyperkinetic Scale ($p < .07$). A cursory power analysis showed that if the originally targeted 18 vestibular-alone subjects showed the same trend, the difference would have been significant ($p < .05$).

The trend of vestibular stimulation-alone advantage in the vestibular-alone series contrasts sharply with the more random-appearing data in the visual-alone series. The
The effect size could not be directly calculated by regression to the mean, statistical regression was controlled by calculating improvement from the baseline 2 score. History also does not seem an adequate explanation, because the amount of change from baseline 1 to baseline 2 did not significantly correlate with the length of time elapsing between baseline 1 and baseline 2. Furthermore, the bulk of improvement on teachers' ratings occurred during the 12 weeks of treatment; there was relatively little further improvement during the year from the end of treatment to follow-up.

It would be unusual to find a Hawthorne or placebo effect persisting for longer than one year. Most treatment of persons having ADD with HA shows a breakdown of placebo effect within a few weeks. If most of the improvement shown here was due to Hawthorne effect, then it is a special kind of Hawthorne effect, which appears to have possibilities as a therapeutic tool. Perhaps the one-year lasting effect is explained by a virtuous cycle or a better level of psychosocial homeostasis initiated by Hawthorne effect at the time of treatment. However, the most parsimonious explanation would seem to be the therapeutic effectiveness of the treatment.

Figure 5
Conners hyperkinetic factor subscale mean sums for 18 subjects in visual-alone series

The sum of changes from baseline 2 (2nd baseline) to the end of all 12 wks of treatment was significant (p < .001). Lower score is better; the mean ratings shown for the fourth week of each 4-week condition in the Latin square crossover were obtained by subtracting from 2nd baseline the mean adjusted change during that condition (shown and explained in Table 2).

Visual-only condition showed just slight advantage over combined or TAV conditions. The contrast is demonstrated in the estimated effect sizes (5) in Table 2. Effect size [f, as defined by Cohen (5, p 274)] is the ratio of the standard deviation among means to the average standard deviation within means. An effect size of .5 (large) indicates that 20% of the total variance is due to differences between the means. An effect size of .1 (small) indicates that only 1% of the total variance is due to differences between means (5). Thus, on the Conners scale sums, 20% of the total variance in the vestibular-alone series was due to differences between treatment conditions, whereas only 1% of the variance was due to between-treatment differences in the visual-alone series.

For each series the carryover effect size was roughly the same magnitude as the treatment effect. (Both effects were larger for the vestibular-alone series.) Sequence effect size could not be directly calculated. However, the size of the mean square associated with sequence effect was the same for both series and was roughly the same as the mean square associated with treatment and carryover effects for the visual-alone series. Of course, this was considerably smaller than the mean square associated with treatment and carryover effects in the vestibular-alone series.

Discussion

Impressive improvement in behavior ratings occurred between baseline 2 and the end of treatment. This improvement was sustained at follow-up more than one year later. The amount of improvement from baseline 2 to follow-up was significant, both statistically and clinically. It was about two-thirds of the amount of improvement found with short-term administration of stimulant drugs (6, 7). The improvement seen in this study is not easily explained by history or statistical regression to the mean. Statistical regression was controlled by calculating improvement from the baseline 2 score. History also does not seem an adequate explanation, because the amount of change from baseline 1 to baseline 2 did not significantly correlate with the length of time elapsing between baseline 1 and baseline 2. Furthermore, the bulk of improvement on teachers' ratings occurred during the 12 weeks of treatment; there was relatively little further improvement during the year from the end of treatment to follow-up.

For each series the carryover effect size was roughly the same magnitude as the treatment effect. (Both effects were larger for the vestibular-alone series.) Sequence effect size could not be directly calculated. However, the size of the mean square associated with sequence effect was the same for both series and was roughly the same as the mean square associated with treatment and carryover effects for the visual-alone series. Of course, this was considerably smaller than the mean square associated with treatment and carryover effects in the vestibular-alone series.
alone condition seemed as effective as the combined condition. Could it be that with visual rotation alone, the visual input excites the vestibular system as much as with combined stimulation? Or could the combined condition actually be a visual condition with the visual input completely overriding the vestibular input?

These speculations must be tempered by the realization that despite the impressive effect sizes, the traditional p values comparing treatment conditions did not reach the significance levels of the previous study (1). This could be partially explained by the unfortunate loss of one iteration of subjects in the Latin square for the vestibular alone series: 18 subjects, rather than 12 with the same trend, would have reached significance on at least one of the two scales. Another possible explanation could be that this nonclinical outreach sample responded differently from the previous clinical sample. The most intriguing possibility follows.

What if the control condition were not really a sham but merely a less effective treatment? Although the plot of the Latin square crossover in Figures 1-3 implies worse scores on the TAV control condition than at baseline, this is an artifact of the calculation: the mean change during the condition (see Table 2) was subtracted from baseline. Two-thirds of these changes during TAV were from a preceding rotatory-stimulation score (see Table 1) and showed deterioration, which, by averaging, masked the improvement found one-third of the time when TAV immediately followed baseline. The raw scores for the TAV condition were better than baseline scores, although not by as much as for rotatory stimulation. In each of the TAV control perceptual experiences, the subject concentrated on a single sensory input at one time: tactile (subject received puff of air on skin), auditory (subject listened to the optokinetic drum rotate with no visual or kinesthetic input), or visual. This was arranged deliberately to make the control condition analogous to the experimental treatments and thus produce a convincing sham. Such unimodal structured sensory experiences conceivably could benefit a child in three ways: a) it provides practice in paying attention for a given time, by reinforcing this behavior with the attention of a liked adult, b) concentration on a single perceptual modality at a time may have a beneficial effect on cerebellar or other processing [support for this theory can be found in the significant academic and behavioral benefit from channel-specific perceptual stimulation reported elsewhere (7, 8), and c) concentration on the nondescript sound of the optokinetic drum seems to have a trance-inducing effect. There is a possibility that the children were inadvertently being taught autogenic relaxation, which has been claimed useful for such children by a few authors (9, 10).

One way to test whether single-channel sensory attention is calming is to use as a control condition a simultaneously multisensory experience, a control condition that we wish, in retrospect, had replaced the successively TAV condition. Another way would be to have a control that does not emphasize the sensory-perceptual area; however, devising a convincing sham of that type would be quite a challenge.
importance of history than did the correlation of change in baselines with the time elapsed between baselines. We also could have added another contact control group that merely had an opportunity to play for 10 or 15 minutes, using the student assistant as a further measure of Hawthorne effect; however, this would have been similar to the control condition of the original study, which we were attempting to improve.

If rotational vestibular stimulation is effective for ADD and HA, our data do not support the theory that the mechanism of benefit is through improvement of visual-motor function, as measured by the Bender-Gestalt Kopitz score. The lack of correlation with Kopitz scores is neither evidence for nor against the interpretation of the behavioral improvement as a Hawthorne effect. While dextroamphetamine produces both behavioral and Kopitz score improvement, levodopaamphetamine produces comparable behavioral improvement and does not change the Kopitz score (11).

While this study does not definitively establish that rotational vestibular stimulation benefits selected cases of ADD with HA, it does suggest that if the stimulation is effective, then the vestibular stimulation is more effective than the visual.

ACKNOWLEDGMENTS

The authors thank the students from the Southwestern City Schools (Franklin County), the Grandview Heights City Schools, and the Upper Arlington City Schools for their cooperation and Patrick McNamara, Dr. Elizabeth Reiss, and Dorothy Walker for their assistance. This study was supported by The Ohio Department of Mental Health and by The Ohio State University College of Medicine.

REFERENCES


The American Journal of Occupational Therapy 91