Improving Attention and Preventing Reading Difficulties among Low-Income First-Graders: A Randomized Study

Eric Dion · Catherine Roux · Danika Landry · Douglas Fuchs · Joseph Wehby · Véronique Dupéré

Published online: 22 June 2010 © Society for Prevention Research 2010

Abstract Students’ inattention is predictive of reading problems and of non-response to effective reading intervention. In this randomized study, 58 first-grade classrooms located in 30 schools were assigned to a control condition or to one of two intervention conditions. In these last two conditions, peer-tutoring activities were conducted to improve classroom reading instruction. In one of the intervention conditions, the Good Behavior Game was also implemented to maximize students’ attention during reading lessons. Both interventions were effective: peer-tutoring activities helped students improve their reading skills and attention was generally higher when the Good Behavior Game was implemented. Contrary to expectations however, students identified as inattentive at pretest did not become better readers when the two interventions were implemented.

Keywords Attention · Reading problems · Classroom instruction · Peer-tutoring · Low-income

Reading problems among young low-income students probably result from multiple factors, including under-stimulation at home, inconsistent school attendance and poor instruction (Torgesen 2002). Among predictive factors of reading problems, one of the strongest is teacher-rated classroom inattention, even in the absence of clinical-level attention-deficit hyperactivity disorder (e.g., Dally 2006). Furthermore, inattentive students’ reading problems are resistant to prevention efforts; that is, current research-based reading interventions are not effective with many of these students (Al Otaiba and Fuchs 2006; Rabiner et al. 2004). It thus seems critical to address low-income student inattention in class.

Classroom Interventions to Increase Students’ Attention

At least two interventions can be used to increase students’ attention in class. The first is self-monitoring. To illustrate, Rooney et al. (1984) trained a teacher to implement this intervention with a group of second-graders exhibiting widespread attention problems. During seatwork, a tape-recorded signal was presented at random intervals. Upon hearing the signal, students recorded whether they were concentrating on their work and resumed working. Student attention rose by 76%. Edwards et al. (1995), for their part, trained three students with attention-deficit hyperactivity disorder to self-monitor. Following the introduction of the intervention, all students increased their attention and correctly answered a greater number of reading comprehension probes, a result suggesting that improving classroom attention has a positive effect on reading skills.

This research was supported by the Social Sciences and Humanities Research Council (Grant No. 410-2005-1474), with additional support by the Fonds québécois de la recherche sur la société et la culture (Grant No. 96785). We gratefully acknowledge the contribution of Commission scolaire de Montréal administrators, principals and teachers who made this study possible, especially Lyne Leblanc, Marielle Cardinale, Josée Pethel, Chantal Legault, Maryse Gagnon, Nathalie Morin and Diane Talbot.

E. Dion (✉) · C. Roux · D. Landry
Département d’éducation et formation spécialisées, Université du Québec à Montréal, C.P. 8888, Succ. Centre-ville, Montreal, Quebec H3C 3P8, Canada
e-mail: dion.e@uqam.ca

D. Fuchs · J. Wehby
Department of Special Education, Vanderbilt University, Nashville, TN, USA

V. Dupéré
Department of Child Development, Tufts University, Medford, MA, USA

© Springer
Reid and Harris (1993), however, found that the intervention had positive effects on students’ (N=28) attention, but not on their spelling skills. In another study, self-monitoring improved spelling skills of only two out of four participants with attention-deficit hyperactivity disorder (Harris et al. 1994). It is therefore unclear whether self-monitoring improves young students’ academic skills, including reading (for a review, see Harris et al. 2004). Furthermore, the frequent audio signals used to cue students’ self-monitoring interfere with teachers’ presentations (Glynn and Thomas 1974). Because of this, the intervention cannot be implemented during whole-class lessons, a format that accounts for about half the time devoted to reading instruction in first-grade classrooms (National Institute of Child Health and Human Development Early Child Care Research Network 2002).

A second intervention, the Good Behavior Game (GBG), can be implemented both during teacher-directed whole class lessons and independent seatwork (Barrish et al. 1969). A group contingency is used to reduce students’ disruptions and increase their attention: The classroom is divided in two teams and the entire team is penalized when one of its members disrupts ongoing activities. During whole-class lessons, teachers occasionally interrupt their presentation to signal to teams that they are losing points, but otherwise the lesson proceeds normally. GBG effectiveness in reducing disruptions has been demonstrated in numerous single-case studies (for a review, see Dion et al. 2009). In most of these studies, student learning was not assessed. GBG usefulness in creating environments more conducive to learning is nevertheless apparent. For instance, in Medland and Stachnik’s (1972) case study, a fifth-grade teacher was able to cover 25% more material after the GBG was introduced.

Two large-scale randomized studies have examined the effect of the GBG on student achievement. In one, first-grade classrooms were assigned to a control condition or a GBG condition (Dolan et al. 1993). Compared to controls, GBG students were less disruptive (and presumably more attentive) at the end of the school year, but the GBG had no effect on standardized reading scores. Although these results suggest that the GBG does not by itself help young students learn to read more easily, the researchers reasoned that it could do so if reading instruction was also improved. In their second study, first-grade classrooms were assigned to different conditions, two of which are relevant here: a control condition and a classroom-based intervention condition (Ialongo et al. 1999). In this second condition, teachers were trained to implement the GBG, but also to use new reading and mathematics materials of an unspecified nature. At the end of first grade, students in the intervention condition were rated as being more attentive. The intervention also had a small effect on standardized reading scores, but only for boys entering first grade with above-average reading skills. These inconsistent results could be attributable to the use of instruction materials of undocumented effectiveness.

The Present Study

Why could it be important to improve both classroom attention and reading instruction? Interventions that target classroom inattention are not instructional strategies per se, and it has been argued that increased attention will facilitate learning only when students can understand instruction (Harris et al. 1994). Of course, the more complex the task, the more likely it is that students will not understand instruction, and learning to read is clearly a complex task for low-income students (Dion et al. 2004). Unfortunately, the effectiveness of a multi-component intervention that effectively targets both attentional and instructional factors has not been empirically tested.

In this study, we compared the effectiveness of improved reading instruction alone or in combination with an intervention designed to increase student attention. To improve reading instruction, teachers implemented evidence-based peer-tutoring activities. To increase student attention, they used an adaptation of the GBG. Teachers and groups of first-graders were assigned to one of three conditions: a control condition, a peer-tutoring only condition, or a combined condition (peer-tutoring and GBG). In a given class, all students participated in the same activities, but only a few were assessed. The assessed sample included inattentive students at-risk of reading problems, along with some of their less at-risk peers.

Our goal was to examine how student attention moderates the effectiveness of the interventions. We expected that, when combined, peer-tutoring activities and the GBG would exert a synergistic effect on inattentive students’ reading achievement. In other words, we hypothesized that inattentive students would learn to read more easily when the two interventions were implemented. By contrast, it was expected that peer-tutoring activities would help attentive students learn to read, whether the GBG was also implemented or not.

Method

Participants

Teachers Fifty-eight first-grade teachers (93% female) participated in this study. These teachers worked in 30 francophone schools located in some of Montreal’s poorest neighborhoods. On average, 50.4% (SD=9.5) of households in these neighborhoods reported a yearly income
below the Canadian poverty line. Teachers were not paid for their participation.

**Students** Student selection was based on a measure of rapid letter naming (RLN, see Instruments) and teacher’s nominations for inattention (“Who are the most inattentive students?”), both obtained in September for all students for whom parental consent was available. A two-stage selection procedure was used. First, we identified the four students most at-risk of reading problems based on their low RLN scores, along with the three students whose RLN scores were the closest to the classroom average (e.g., Mathes et al. 1998). Second, if necessary, substitutions were made to ensure that all inattentive students with eligible RLN scores (i.e., low or average scores) were included. The sample ($N=409$, 44% female) thus included an overrepresentation of inattentive students at-risk of reading problems (31.3%, $n=128$) because these students were of prime interest, and because it was incorrectly assumed that many of them would be lost due to family residential instability over the course of the school year (a negligible 2.4% of attrition was observed for the sample as a whole).

**Experimental Design**

Groups of three similar schools were created by considering the percentage of households with a yearly income below the poverty line. In each of these groups, schools (i.e., all participating groups and teachers) were randomly assigned to one of three conditions. In the control condition, teachers continued to teach reading and manage their classroom in their usual fashion. They received the training and intervention materials at the end of the school year. In the peer-tutoring only condition, teachers were trained to implement peer-mediated reading activities but continued to manage student behavior as they normally did. In the combined condition, teachers were trained to implement both the peer-tutoring activities and an adaptation of the GBG to increase student attention during regular reading lessons.

Originally, 60 teachers had agreed to participate and 20 had been assigned to each condition, but 1 teacher fell ill in the middle of the school year and her substitute declined to participate. A second teacher was eliminated because of her repeated refusals to allow progress assessments of her students. Thus, the study was based on 20 control teachers, 20 peer-tutoring only teachers, and 18 combined teachers.

**Interventions**

**Peer-Tutoring Reading Activities** These activities are *Apprendre à lire à deux* (Dion et al. 2005a), a French adaptation of the First-grade Reading Peer-Assisted Learning Strategies (PALS, Mathes et al. 1998). In the peer-tutoring only and combined conditions, teachers conducted peer-tutoring activities with their students for 30 min, three times a week, from mid-October to April, during time allocated to French (i.e., language arts). Teachers in these two conditions were not required to make any other changes to the content of their typical reading program. As in the control classrooms, their reading program had a whole-language orientation with only occasional instruction on letter sounds and word decoding.

For peer-tutoring activities, students were paired so that strong readers partnered with weaker ones. Partners took turns pronouncing letter sounds (3 min), blending sounds to form words (4 min), reading sight words (3 min) and reading a short decodable story (5 min) from a lesson sheet. Starting in February, they also alternated reading illustrated books (10 min). Partners were trained to correct each other’s mistakes. Each class was divided in two teams and partners earned points for their team by completing activities. At the end of the week, the team with the most points was asked by the teacher to stand while the other team clapped. Next, the first team applauded the second team in recognition of their hard work. New pairs and teams were formed every month.

The original version of the peer-tutoring activities has proved effective with different categories of students, including students at risk of reading disabilities and average students (see Fuchs and Fuchs 2005). Based on these results, the U.S. Department of Education’s Institute of Education Sciences has recommend their use (U.S. Department of Education Institute of Education Sciences What Works Clearinghouse, July 2007). Nonetheless, at least one study suggests that PALS activities are not effective with inattentive students (Al Otaiba and Fuchs 2006).

**GBG** Developed for this study, *Attention je lis!* is an adaptation of the GBG explicitly centered on attention (Dion et al. 2005b). From November to April, this intervention was implemented daily, for a 15 min period, during what combined condition teachers considered their most important reading lesson of the day, excluding peer-tutoring activities. This lesson was typically conducted in a whole-class format. The adapted GBG was implemented during regular reading lessons rather than peer-tutoring activities, because these last activities usually generates a high level of attention (DuPaul et al. 1998). Although they were conducted at different times, the GBG and peer-tutoring activities relied on the same team competition for motivation.

During a GBG period, teachers presented content in their usual fashion; they only managed their students’ attention differently. They signaled the beginning of the period by placing the intervention logo on the board. From this
moment on, every time students were disruptive, they could cost their team as many as three points and hence lessen its chances of winning the weekly contest. When a disruption occurred, teachers immediately signaled it and subtracted points from the team’s total. In the original version of the GBG (Barrish et al. 1969), the goal for students was simply to avoid losing points by behaving appropriately. In our adapted version, student attention was also proactively reinforced. At the end of the 15 min period, teachers commented on the teams’ performance and assigned a maximum of five points if they judged a team to have been optimally attentive. Teachers concluded the period by reinforcing verbally one or two students who had shown exemplary behavior.

**Instruments**

*RLN* RLN was used to select students for assessments. It also serves as a pretest variable. RLN assesses the number of letter names a student identifies in 1 min. Four columns of letters are displayed on a sheet, in random order. When students seemed stuck for more than 3 s, they were told to move to the next letter. Among beginning readers, performance on RLN is an excellent predictor of reading skills acquisition (e.g., Ritchey and Speece 2006).

**Teacher-Rated Inattention** Teachers rated students’ pretest classroom behavior on the short version French adaptation of the *Conners’ Teachers Rating Scale* (Conners 2000). Items were rated on a four-point scale (“Not at all true” to “Very true”). Because of previous research suggesting that lack of response to reading instruction is specifically associated with attention problems (Rabiner et al. 2004), we only used responses to five non-redundant items describing inattentive behaviors (e.g., “Inattentive, easily distracted,” “Daydreams”). Response to these five items were summed to create an attention subscale ($\alpha = .91$), with higher scores indicating greater attention problems.

**Observed Classroom Attention** Observations were conducted to provide a measure sensitive to intervention effect on attention (see Stoolmiller et al. 2000). Each student was observed during a single whole-class reading lesson. In combined condition classrooms, observations were conducted while teachers implemented the GBG. In the control and peer-tutoring only conditions, observations were made while teachers presented a lesson and managed the group’s attention in their usual fashion. A time sampling procedure was used. Students were observed for rounds of 12 consecutive 5 s intervals, until a total of 2.5 min of observation had been collected for each student. During each 5 s interval, students’ behavior was coded as optimally attentive, disruptive or inattentive. To be considered optimally attentive, students had to be, during the whole interval, correctly seated and oriented toward the relevant stimuli. They also had to abstain from breaking rules and making unnecessary movements. Students were considered disruptive during an interval if, at one point or another, they emitted noises, moved in ways that interfered with ongoing activities or did not follow rules. They were coded as inattentive when the two previous codes did not apply (i.e., when they were less than optimally attentive but not disruptive). A second observer was present for 15% of these observations. Correlations between proportions of behaviors for the two observers ($r=.88–.95$) indicate an adequate inter-rater agreement. Scores were calculated as the proportion of intervals during which the behavior was observed on the total of intervals observed for the student.

**Progress Assessments** From October through April, recognition of grapheme-phoneme correspondence was assessed every other week. This measure is fluency-based: Students had 1 min to recognize as many stimuli as they could and scores correspond to the number of correctly pronounced stimuli per min. Different lists of 100 randomly sampled stimuli were used for each assessment. Specifically, students were presented commonly-used words with, underlined in each, a letter or group of letters representing a single phoneme (e.g., “match”). Students were asked to pronounce only the underlined grapheme, as it should be pronounced in the word. For instance, that they had to correctly identify “s” pronounced as /z/ rather than /s/.

**End-of-Year Assessments** At the end of April, word recognition skills were assessed with the *Échelles de compétences en lecture* (Desrochers 2008). Students were asked to read orally a list of words and a list of non-words. Students had 45 s to read each list. Two scores were calculated: the number of correctly read words and non-words. Test-retest reliability is strong ($r=.90–.96$) for both scales. Reading comprehension was assessed with *Lire pour comprendre*, a test modeled after the *Comprehensive Reading Assessment Battery* (Fuchs et al. 1988). Students were permitted 3 min to read orally a brief (114 words) narrative text. They were then asked to answer six open-ended questions. Answers were rated as incorrect, partially correct, or correct. Inter-rater agreement, calculated for 24% (100/409) of the sample, was strong (92–99%). The comprehension score ($\alpha = .77$) corresponds to the number of questions answered correctly.

**Fidelity of Implementation** The fidelity with which peer-tutoring activities and the GBG were implemented was observed over one (peer-tutoring only condition) or two (combined condition) sessions. To establish inter-rater agreement, a second observer was present for a number of
randomly selected sessions. For peer-tutoring activities, fidelity was observed in February with a translation of the detailed checklist developed for the First-Grade Reading PALS (Mathes et al. 1998). Both teachers’ (29 items) and students’ (71 items) behaviors were observed because students had to autonomously implement certain aspects of the intervention (e.g., listening to their partner). Each item was scored as occurring, not occurring, or not applicable and the percentage of items scored as occurring with respect to the number of items scored as either occurring or not occurring was calculated. Inter-rater agreement, calculated for 26% (10/38) of intervention classrooms, corresponds to the percentage of items scored similarly by both observers. Mean agreement is 91.7% (SD=3.6). For the GBG, fidelity was observed in March with a brief checklist developed for this study. Once again, items were scored as occurring, not occurring, or not applicable. The checklist is short (11 items) because this intervention is relatively simple and because only teachers’ behaviors were scored. No aspects of the intervention were implemented autonomously by students and implementation could be considered strong even if students were inattentive and disruptive. Mean agreement, calculated over 28% (5/18) combined condition classrooms, is 94.2% (SD=8.8).

We were also interested in documenting how often teachers had implemented interventions. There was a simple mean to do this for peer-tutoring activities. Teachers used one consecutively numbered lesson sheet for each session of peer-tutoring activities. At the end of the intervention period, research assistants visited classrooms and noted which lesson sheet the group was using. By dividing the number of lesson sheets the group had completed by the duration of the intervention in weeks, we obtained an estimate of the average number of peer-tutoring sessions conducted weekly.

Procedure

Student Selection and Assessment Students were selected for assessment around mid-September. The RLN was administered, teachers nominated inattentive students and completed the Conner’s Rating Scale for all selected students. Reading assessments, including the RLN, were administered on a one-to-one basis, in a quiet room close to the classroom. Because examiner unfamiliarity may depress the performance of young low-income students (e.g., Fuchs and Fuchs 1986), each student was assessed by only one research assistant. This research assistant also conducted classroom observations (February). To avoid introducing bias, research assistants were trained to conduct assessments and observations in a standard manner.

Training Around mid-October, the principal investigator offered a half-day workshop to groups of teachers that had been assigned to the peer-tutoring only condition or to the combined condition. The first two hours of the workshop were dedicated to the peer-tutoring activities. An overview of the intervention was presented and questions related to logistics, pairing of students and commonly occurring problems were covered. Teachers were then paired to role-play all the peer-tutoring activities. Next, they received a detailed manual that includes lesson plans to facilitate student training (Dion et al. 2005a). Teachers were asked to train their group to conduct peer-tutoring activities during the next two weeks.

The last hour of the workshop was dedicated to the GBG (only teachers in the combined condition attended). A rational for this intervention was presented with an emphasis on the importance of students’ attention during reading lessons. The principal investigator explained how students could make their team win or lose points. He asked teachers to implement the GBG each day for 15 min, but not more. Detailed lesson plans to facilitate student training were included in the teacher’s manual (Dion et al. 2005b). Combined condition teachers were asked to introduce the GBG immediately after having completed the two-week training for GBG, toward the beginning of November.

In Class Support and Testing The research assistant conducting assessments was available to offer in-class support on an as needed basis. Support was offered either because it was directly requested or, occasionally, because discussion with teachers or classroom visits suggested a less than optimal implementation.

Results

Fidelity of Implementation

Peer-tutoring activities were implemented in classrooms assigned to both the peer-tutoring only and combined conditions. In the peer-tutoring only condition, observed fidelity of implementation was 92% (SD=11) for teachers and 88% (SD=6) for students. In the combined condition, it was 95% (SD=6) for teachers and 90% (SD=8) for students. The frequency at which these activities were conducted was also estimated based on the number of lesson sheets completed over the intervention period. In the peer-tutoring only condition, 3.0 (SD=0.4) sessions were conducted each week, compared with 2.9 (SD=0.5) in the combined condition. In other words, teachers in both conditions conducted peer-tutoring activities as planned and implemented these activities three times a week. The
other intervention, the GBG, was implemented only in the combined condition. For this last intervention, teachers’ observed fidelity of implementation was also strong, 92% (SD=8).

Pretest

Schools rather than individual students were assigned to conditions. To take this source of statistical dependency into account, a multilevel approach was implemented with the Hierarchical Linear Models (HLM) software (Raudenbush et al. 2004). Non-standardized beta coefficients are reported.

Two-level models (students within schools) were used to compare (level one) student pretest scores as a function of (level two) school condition. The three conditions were represented by two dichotomous variables: a first variable indicating the presence of peer-tutoring activities (peer-tutoring only and combined condition) and a second variable indicating the presence of the GBG (combined condition).

Means of teacher-rated attention were virtually identical in the three conditions. However, compared with their counterparts in the control and combined conditions, students in the peer-tutoring only condition had slightly higher RLN scores (Table 1), a difference that almost reached statistical significance, $B=4.70$, $t(df=27) = 1.96$, $p=.060$. To control for possible differences between conditions, RLN is included as a covariate in all analyses of reading scores.

### Table 1 Pretest, mid-year, and end-of-year scores (means, standard deviations) by condition

<table>
<thead>
<tr>
<th>Time administered/measure</th>
<th>Control</th>
<th>PTO</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pretest</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher-rated attention</td>
<td>11.3</td>
<td>11.6</td>
<td>11.5</td>
</tr>
<tr>
<td></td>
<td>(4.5)</td>
<td>(4.6)</td>
<td>(3.9)</td>
</tr>
<tr>
<td>RLN</td>
<td>22.1</td>
<td>26.7</td>
<td>22.2</td>
</tr>
<tr>
<td></td>
<td>(10.6)</td>
<td>(13.2)</td>
<td>(11.9)</td>
</tr>
<tr>
<td><strong>Mid-year</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed optimal attention (proportion)</td>
<td>0.65</td>
<td>0.69</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>(0.25)</td>
<td>(0.22)</td>
<td>(0.16)</td>
</tr>
<tr>
<td><strong>End-of-year</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word recognition</td>
<td>17.0</td>
<td>20.3</td>
<td>21.0</td>
</tr>
<tr>
<td></td>
<td>(11.7)</td>
<td>(12.0)</td>
<td>(11.8)</td>
</tr>
<tr>
<td>Non-word recognition</td>
<td>11.8</td>
<td>14.5</td>
<td>14.9</td>
</tr>
<tr>
<td></td>
<td>(9.6)</td>
<td>(9.5)</td>
<td>(9.8)</td>
</tr>
<tr>
<td>Comprehension</td>
<td>1.9</td>
<td>2.5</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td>(1.6)</td>
<td>(1.7)</td>
<td>(1.7)</td>
</tr>
</tbody>
</table>

Means and standard deviations are calculated using individual scores. *PTO* peer-tutoring only, *RLN* rapid letter naming

Classroom Attention

Observations of student attention were conducted during a regular reading lesson. Teachers managed this lesson in their usual fashion (control and peer-tutoring only condition) or with the GBG (combined condition). Observations thus indicate the effectiveness of the GBG in improving student attention. Probably due to the short duration of the observation period for each student, disruptive behaviors were rarely observed. Accordingly, analyses focus on proportions of optimal attention during regular reading activities (Table 1).

As with the reading outcomes (see below), we were interested in the moderating role of pretest (teacher-rated) attention on intervention effectiveness. To this end, cross-level interactions were examined (Raudenbush and Bryk 2002). Each student’s pretest attention score was multiplied by the value of the dichotomous variables representing the presence of the two interventions in this students’ school. The two products were entered as student-level variables in the regression equations, along with the original variables used in their calculation. A significant parameter associated with a dichotomous variable indicates a generalized (average) effect of the intervention. A significant parameter associated with a product indicates that pretest attention moderated the effectiveness of the intervention.

Peer-tutoring did not have an effect on attention during regular reading lessons, either in a generalized manner, $B=0.05$, $t(df=27) = 0.60$, n.s., or as a function of pretest attention, $B=-0.00$, $t(df=369) = -0.50$, n.s. As expected however, the GBG had a generalized effect, $B=0.17$, $t(df=27) = 3.03$, $p<.01$, with a marginally significant moderating role of pretest attention, $B=0.01$, $t(df=369) = 1.82$, $p=.068$. To interpret this effect, expected levels of optimal attention during regular reading lessons were estimated for students with pretest attention one standard deviation below or above the sample mean. Effect sizes were calculated as the arithmetic difference between estimated values divided by the residuals standard deviation for the outcome (Stoolmiller et al. 2000). In classrooms where the GBG had not been implemented, attentive students were estimated to be optimally attentive 71% of the time and inattentive students only 62% of the time. In classrooms where the GBG had been implemented, estimated values for optimal attention were high for both attentive (84%) and inattentive (82%) students. The effect of the GBG on optimal attention was large for attentive students (0.81) and very large for their inattentive classmates (1.22).

Reading Progress

Three-level growth curve models were used to analyze bi-monthly progress assessments. Measurement occasion
(level one) was considered nested within students (level two), and students nested within schools (level three). To model both linear (uniform) and curvilinear (accelerating or decelerating) rates of progress, time was expressed in number of months and squared number of months elapsed since the pretest. Three random parameters describing progress were initially estimated for each student: an estimated pretest score (an intercept), a linear rate of progress and an accelerating or decelerating rate of progress. The model included pretest RLN as a (level two) control on these parameters.

Linear rates of progress for grapheme-phoneme correspondences were positive and significant, $B=7.37$, $t(df=29)=15.27, p<.001$, but there was also a significant deceleration, $B=-0.50$, $t(df=29)=-9.37, p<.001$. In other words, progresses were initially rapid but tended to slow down toward the end of the intervention period. Because the deceleration parameter did not vary as a function of condition, it was fixed at its average value for all students.

The number of statistical comparisons is thus reduced because intervention effects are expressed solely in terms of between-schools differences in linear progress rates. There was significant between-school variance on both progress and pretest attention, $B=0.52$, $t(df=27)=3.31, p<.01$, but less so for inattentive students, $B=0.18$, $t(df=393)=-2.67, p<.01$. Based on estimated mean linear rates of progress and their standard deviation, the effect size of the peer-tutoring activities was very large for attentive students (1.04) but only moderately large for inattentive students (0.40). Contrary to expectation, there was no indication that the GBG influenced the linear rates of progress either in a generalized manner, $B=0.52$, $t(df=27)=0.87, n.s.$, or as a function of pretest attention, $B=-0.01$, $t(df=393)=-0.20, n.s.$

**End-of-Year Reading Skills**

End-of-year students’ reading skills (Table 1) were analyzed using two-level models (students within schools). Neither peer-tutoring activities nor the GBG influenced word recognition. The average effect of peer-tutoring activities on non-word recognition was not significant, $B=1.54$, $t(df=27)=0.75, n.s.$, but there was a marginally significant moderating effect of pretest attention for this outcome, $B=-0.38$, $t(df=392)=1.78, p=.078$. Similarly, there was no average effect of peer-tutoring activities on comprehension, $B=0.52$, $t(df=27)=1.53, n.s.$, but a moderating role for pretest attention that in this case reached significance, $B=-0.09$, $t(df=392)=-2.31, p<.05$. Once again, the GBG had no significant effect on non-word recognition or comprehension.

To interpret the effect of peer-tutoring activities, expected values for non-word recognition and comprehension were estimated for students with pretest attention one standard deviation below or above the sample mean. Once again, effect sizes were calculated by dividing the arithmetic difference between estimated values by the residuals standard deviation for the outcome. For attentive students, peer-tutoring activities had a moderately large effect on non-word recognition (0.60) and comprehension (0.76). For inattentive students, the corresponding effect size was negligible for non-word recognition (0.10) and substantially smaller for comprehension (0.25).

**Non-Response**

Typically, there are wide variations in the degree to which individual students benefit or respond to an intervention (Dion et al. 2004) and linear comparisons offer somewhat limited insights into the effectiveness of an intervention. To get a more realistic idea of its strengths and limitations, it is also important to estimate the proportion of students that do not respond to the intervention.

The bi-monthly grapheme-phoneme correspondence assessments were used to identify non-responders (i.e., students who were not learning to read). Empirical Bayes estimates of individual intercepts and linear slopes were estimated by running a three-level growth curve model with only two predictors: number of months and square number of months elapsed since the pretest. In line with previous analyses, the deceleration rate was constrained at its average value for all students.

Following McMaster et al. (2005), non-responders were identified by considering individual students’ (linear) rate of progress and level at the end of the intervention period (i.e., 4 months after the pretest). To be identified as a non-responder, a student had to have both a rate of progress and a level in the lowest third of the sample. In other words, non-responders were students who did not make enough progress to attain a satisfactory performance level. Using these criteria, 28.9% ($n=115$) of the students were categorized as non-responders (the other students were categorized as responders). The fact that non-responders had problems learning to read is made clear by comparing this last group with responders. The former had much lower pretest (RLN) and end-of-year scores (word and non-word recognition and comprehension) than the latter, with effect sizes between .71 and 1.28.

There were 47.1% non-responders in the control condition, compared with only 19.8% and 18.1% in the peer-tutoring...
only and combined condition, respectively. A two-level model (students nested within schools) for dichotomous outcomes was run to determine if these percentages differed by condition and pretest attention level. Odds ratios (OR) with 95% confidence intervals (CI) are reported. Peer-tutoring activities had an average (generalized) effect, OR=0.20, CI=0.05–0.80, p<.05, that was moderated by pretest attention OR=1.22, CI=1.04–1.43, p<.01, but none of the parameters associated with the GBG approached significance. To interpret significant effects, the percentage of non-responders was estimated for students one standard deviation above or below the mean on pretest attention. In the control condition, the percentage of non-responders was high among both attentive and inattentive students (48.6% and 48.9% respectively). In schools where peer-tutoring activities had been implemented, the percentage of non-responders was much lower, especially among attentive students (5.5%), but also among inattentive students (26.5%).

Discussion

In this study, intervention teachers implemented a reading intervention (peer-tutoring activities) alone or in combination with the GBG. They conducted the interventions with adequate fidelity and significant positive effects were observed on reading skills and attention during lessons. As expected, inattentive students benefited less from peer-tutoring activities than their more attentive classmates. Contrary to expectations, however, these students did not become better readers when teachers implemented the GBG in addition to peer-tutoring activities. In other words, concurrently improving reading instruction and increasing attention in class did not have a synergistic effect on inattentive students’ reading skills. These results are discussed in detail.

In previous studies suggesting that increased attention did not facilitate learning, instruction has not been meaningfully improved (e.g., Ialongo et al. 1999) and could have been of less than optimal quality, explaining why students did not benefit from being more attentive. We avoided this problem by asking teachers to implement evidence-based peer-tutoring activities. Indeed, our results reaffirm the effectiveness of the GBG on reading skills, but effect sizes were smaller than for their attentive classmates. Similarly, inattentiveness quintupled (26.5/5.5) the risk of being a non-responder to peer-tutoring activities (see also Al Otaiba and Fuchs 2006). Interestingly, the correlation between inattention in class and non-response is also observed when the intervention is offered one-on-one by a trained adult outside the classroom (Al Otaiba and Fuchs 2002), a context where attention is maximized. It thus appears that reading interventions, no matter how intensive or engaging, cannot by themselves compensate for the lost classroom learning opportunities caused by inattention. It also leads to think that increasing attention would help inattentive students learn to read if instruction was effective.

This, however, is not what we observed. Although attention in class seems critical, inattentive students did not become better readers in schools where our adaptation of the GBG was implemented, despite its effectiveness when it comes to increasing attention. During the daily 15 min GBG-managed lesson, both attentive and inattentive students were observed to be highly focused on the lesson content. Why is it that increased attention failed to improve reading outcomes? A more intensive implementation of the GBG (e.g., 3 hours a day, Dolan et al. 1993) would perhaps have been necessary. However, implementing our adapted version of the GBG is demanding for students and teachers, thus limiting possibilities to implement it for extended periods. With this version, students must not only refrain from being disruptive, they have to remain optimally attentive. Pilot testing indicated that it would not have been feasible to implement our adapted GBG for longer periods because remaining concentrated appears quite effortful for many young students. Also, teachers rapidly tire of simultaneously presenting a lesson and keeping track of students’ moments of inattention.

There is at least one alternative explanation for the lack of effect of the adapted GBG on reading skills. Individual characteristics other than inattention could interfere with inattentive students’ ability to learn to read. Inattention is notably associated with a limited vocabulary (Loe et al. 2008) and working memory deficits (Martinussen et al. 2005), and these factors may need to be addressed through a multi-component intervention. Of course, implementing this type of intervention and testing the effectiveness of its different components is challenging. As a preliminary step, it would be useful to carefully examine how inattention and associated factors interact to predict non-response to evidence-based reading instruction. It remains unclear, for instance, whether vocabulary predicts non-response when inattention is controlled (Stage et al. 2003; Torgesen et al. 1999).

The present study is not without limitations. Some effects are marginally significant, and statistical power is an issue. Our sample included a large number of students...
(N=409), but a smaller number of classrooms (N=58) and schools (N=30). In keeping with Cornfield’s (1978, p. 101) oft-cited remark that “Randomization by cluster accompanied by an analysis appropriate to randomization by individual is an exercise in self-deception,” we compared units of randomization (i.e., schools) and statistical power was a function of the number of these units. With 30 schools, the power to detect a medium effect size of .50 at p<.05 was only .29 (Cohen 1988), a noteworthy lack of power for a sample that can be considered large by the standards of school-based intervention research. To alleviate this problem, we could have compared the 58 classrooms rather than the 30 schools. Although unorthodox, this option is appealing statistically and conceptually. Other things being equal, comparing units of interventions (classrooms) rather than units of randomization (schools) could have increased power by a factor of 1.59 (.46/.29). Conceptually, the reading instruction students received is essentially a function of the teacher’s and the group’s characteristics, and this between-classroom variance is consequential for low-income students (Nye et al. 2004). Although preliminary analyses indicate it was not the case here, treating teacher-implemented interventions as school-rather than classroom-level variables may lead to an underestimation of the interventions effectiveness. This issue requires clarification considering the potentially serious consequences of premature abandonment of effective interventions (Stoolmiller et al. 2000).

References


