

■ Follow-up of an Exercise-based Treatment for Children with Reading Difficulties

David Reynolds^{1,*} and Roderick I. Nicolson²

¹*School of Education, University of Exeter, UK*

²*Department of Psychology, University of Sheffield, UK*

This study reports the results of a long-term follow-up of an exercise-based approach to dyslexia-related disorders (Reynolds, Nicolson, & Hambly, *Dyslexia*, 2003; 9(1): 48–71). In the initial study, children at risk of dyslexia were identified in 3 years of a junior school. One half then undertook a 6 month, home-based exercise programme. Evaluation after 6 months indicated that the exercise group improved significantly more than the controls on a range of cognitive and motor skills. Critics had suggested that the improvement might be attributable to artifactual issues including Hawthorne effects; an initial literacy imbalance between the groups; and inclusion of non-dyslexic participants. The present study evaluated the issue of whether the gains were maintained over the following 18 months, and whether they were in some sense artifactual as postulated by critics of the original study. Comparison of (age-adjusted) initial and follow-up performance indicated significant gains in motor skill, speech/language fluency, phonology, and working memory. Both dyslexic and non-dyslexic low achieving children benefited. There was also a highly significant reduction in the incidence of symptoms of inattention. Interestingly there were no significant changes in speeded tests of reading and spelling, but there was a significant improvement in (age-adjusted) reading (NFER). It is concluded that the gains were indeed long-lasting, and that the alternative hypotheses based on potential artifacts were untenable, and that the exercise treatment therefore achieved its applied purpose. Further research is needed to determine the underlying reasons for the benefits. Possible (and potentially synergistic) explanations include: improved cerebellar function (neural level); improved learning ability and/or attentional ability (cognitive level); improved self-esteem and

*Correspondence to: David Reynolds, School of Education and Lifelong Learning, University of Exeter, St. Luke's College, Heavitree Road, Exeter, EX1 2LU, UK; E-mail: David.Reynolds@exeter.ac.uk

self-efficacy (affective level); and improved parental/familial support (social level). Copyright © 2006 John Wiley & Sons, Ltd.

Keywords: ADHD; dyslexia; reading disability; intervention; cerebellum; special educational needs

INTRODUCTION

The major framework for the explanation and treatment of poor reading and dyslexia is in terms of phonological deficit (Snowling, 1987; Stanovich, 1988; Vellutino, Fletcher, Snowling, & Scanlon, 2004). However, it has become clear that phonological deficits are not unique to dyslexia, and in many cases are only one of a larger pattern of difficulties including speed of processing, working memory and motor skill (Miles, 1983; Nicolson & Fawcett, 1995; Ramus *et al.*, 2003; Wolf & Bowers, 1999; Wolff, Michel, Ovrut, & Drake, 1990). Furthermore, while there is no doubt that support for phonological ability is an important component of early literacy support, it is not the only important component, with integrated support for fluency and comprehension needed in addition (Hatcher, Hulme, & Ellis, 1994; NICHD, 2000).

A valuable framework for considering a 'top-down' approach to dyslexia is in terms of cause, symptom and treatment. These three aspects are naturally interdependent, in that when treating a patient with a high fever and headache, it is important to identify the underlying cause—whether it is say malaria, influenza or meningitis—before determining the appropriate treatment. Phonological abnormalities might arise from a number of possible causes—from otitis media to multilingual upbringing to speech disorder to impaired magnocellular function. One underlying explanatory framework that has proved successful in terms of accounting not only for the phonological deficits but also the range and development of problems in dyslexia is in terms of cerebellar deficit (Fawcett, Nicolson, & Dean, 1996; Nicolson, Fawcett, & Dean, 1995, 2001). In brief, the hypothesis proposes that abnormalities in the language-specialist cerebellar circuitry is the underlying cause of the problems in achieving automaticity in language production and reception, and that this leads to the established phonological problems. In many cases, other regions of the cerebellum will also be abnormal, in which case motor skills will also be affected. A key focus of the framework is in terms of learning, because the cerebellum has the circuitry needed to scaffold the development of a range of skills, leading to increases in speed and automaticity. A full description of the theory, the recent explosion of insights from the cognitive neuroscience of the cerebellum, together with a review of criticisms of the theory is provided in Nicolson and Fawcett (2005). A brief analysis is provided in Nicolson and Fawcett (2006).

It is now fully accepted that it is appropriate to consider changes in brain function as reading develops—see Joseph, Noble, and Eden (2001); Salmelin and Helenius (2004) and Shaywitz and Shaywitz (2005) for reviews and Shaywitz *et al.* (2004) for a successful phonological intervention study. Interestingly, almost all studies of brain reorganization in reading have focused on reading or language interventions, including those aimed at magnocellular function (Tallal, Merzenich, Miller, & Jenkins, 1998; Temple *et al.*, 2003). From a theoretical

perspective this makes it difficult to determine whether the effectiveness of the interventions derives from remediation of the presumed underlying cause, or 'merely' from the targetted practice in reading-related sub-skills (Hook, Macaruso, & Jones, 2001).

In a bold and innovative approach, Dore and Rutherford (2001) proposed that, if cerebellar deficit does indeed underlie dyslexia and related learning disabilities, then an intervention designed to improve cerebellar function should lead to benefits even without any specific reading-related support. They therefore constructed an extensive sequence of exercises designed to improve function of the cerebellum and vestibular system, and opened clinics based on the approach. These clinics were originally called dyslexia, dyspraxia and attention-deficit treatment (DDAT) clinics, but are now termed Dore Achievement Centres. It should be stressed that this approach goes beyond the cerebellar deficit hypothesis (which is silent on whether it is possible to improve cerebellar function) to espouse the 'cerebellar treatment hypothesis' (CTH) which claims that it is possible to improve function of the cerebellar/vestibular system, and that this will lead to improved learning ability, which will in turn lead to much improved acquisition of skills taught at school. Literacy will therefore benefit as a side-effect of the cerebellar improvement (assuming that reading is being taught). The treatment is therefore supplementary—intended to supplement whatever reading support is provided—rather than additional (as in most individual literacy support).

In previous research (Reynolds, Nicolson, & Hambly, 2003) we attempted to evaluate the effectiveness of this exercise-based treatment on the performance of the children at risk of dyslexia in a junior school. In the study, the children with highest risk (roughly 13% of the cohort) were divided into two matched groups, and the change in performance over the 6 months study was monitored. One group, the exercise group, undertook the exercise-based treatment daily at home whereas the other group, the control group, undertook no additional activity. School activities were identical between the groups, and indeed the identity of the groups was not disclosed to the teachers. The design was therefore a quasi-experimental 'value-added' one in that it was designed to evaluate the specific contribution of this out-of-school exercise activity.

Cerebellar/vestibular signs were substantially alleviated following the exercise treatment whereas there were no significant changes for the control group. Even after allowing for the passage of time, there were significant improvements for the intervention group in postural stability, dexterity, phonological skill, and (one-tailed) for naming fluency and semantic fluency. Reading fluency showed a highly significant improvement for the intervention group, and nonsense passage reading also improved significantly. Spelling showed no relative change. Significantly greater improvements for the intervention group than the control group occurred for dexterity, reading, verbal fluency and semantic fluency. The improvements in dexterity and postural stability were as expected following a lengthy exercise programme, but the improvements in speed, fluency and phonology suggested a transfer of the training to general cognitive performance. The significant difference in reading suggested a far transfer to some aspects of literacy.

These findings were of considerable potential theoretical and applied interest, in that they appeared to support the CTH. However, the study had

acknowledged limitations, which were also highlighted in a set of invited commentaries (Rack, 2003; Richards *et al.*, 2003; Singleton & Stuart, 2003; Snowling & Hulme, 2003; Stein, 2003). First, the two groups were matched on general risk of dyslexia as assessed by the Dyslexia Screening Test (Fawcett, Nicolson & Dean, 1996) but not explicitly on literacy, in that the range of performance measures under consideration was not confined to literacy. Unfortunately, the exercise group turned out to have a lower initial literacy score, and critics suggested that this might artifactually have permitted greater improvement (alternate hypothesis AH:low-start). Second, given the difference in treatment between the groups, critics suggested that at least part of any relative gains might be attributable to some Hawthorne effect, reflecting transient improvements in self-esteem and motivation rather than any fundamental change (alternate hypothesis AH-Hawthorne). Third, critics noted that the improvements had only been assessed at the immediate post-treatment stage. Most follow-ups of literacy interventions (see NICHD, 2000) find there is a 'wash out' effect in which the initial performance gains are not maintained post-intervention, with the literacy performance slowing converging back onto the original trajectory. The study therefore failed to provide the longer term data needed to properly assess the claim of the CTH that the exercise treatment makes an irreversible change to the child's cognitive architecture, and the alternative hypothesis is that the gains will wash out over time (AH:transient). Fourth, only a minority of the participants had a formal diagnosis of dyslexia, and critics noted that the improvements might in fact derive primarily from the non-dyslexic participants, who might have been underachieving for relatively minor reasons and would therefore get back on track more easily (AH:non-dyslexic).

It is important to note that these criticisms are only alternative hypotheses for explaining the improvements obtained, and there was no direct evidence in favour of any of them. Indeed in subsequent analyses we were able to demonstrate that the between-group differences on literacy were unlikely to have been the cause of the relative improvement (Nicolson & Reynolds, 2003a). It has been widely asserted that the criticisms rebut the study, but this is based on an adversarial 'guilty until proved innocent' philosophy that is the antithesis of the scientific, inquisitorial metatheory. A fairer analysis is that they weaken the strength of the claims, requiring further evidence before the case is convincing (Nicolson & Reynolds, 2003b).

Nicolson and Reynolds (2003b) noted that these issues would be best addressed by means of a carefully designed, large scale, independent investigation. In the continued absence of such a study, valuable evidence may be acquired by monitoring the subsequent progress of the original participants. The present study reports the findings of a 2 year follow-up of the initial study, limited to those children who remained within the same school for that period (the study involved 3 year groups, and the oldest children moved to secondary schools). Immediately after the initial study the original control group were offered the exercise treatment, and all undertook the treatment. The mean duration of the treatment was around 12 months, and so both groups had completed the treatment 6–12 months before the end of the two year follow-up. Consequently the design allows an investigation of all four alternative hypotheses. In order to test these hypotheses we derive differential predictions for the alternative hypotheses as well as the CTH.

Issue 1: Difference between the groups. Hypothesis AH:low-start asserts that group 2 (the original controls) will make relatively little progress, since the improvements of group 1 were artifactual, resulting from their low start. Hypothesis CTH predicts that both groups will make equivalent progress, since the progress is attributable to an improved learning performance for children in both groups.

Issue 2: Reversion to initial performance levels. Hypothesis AH:Hawthorne asserts that the apparent effect of the exercise treatment is attributable to the positive feelings associated with being in a 'special' group and therefore are limited to the time of the exercise treatment. Performance will therefore tail off post-treatment, returning in due course to that predicted without treatment.

Issue 3: Maintenance of treatment gains: Hypothesis AH:transient predicts that (in common with almost all traditional reading-based interventions) there will be a wash-out of the initial performance gains, and that performance will drop back to some level intermediate between pre- and post-treatment levels. CTH predicts, on the contrary, that performance will be maintained, and perhaps continue to improve, since the underlying learning processes have been repaired. Clearly Issue 3 is a more stringent form of Issue 2.

Issue 4: Individual differences: Hypothesis AH:non-dyslexic predicts that only the non-dyslexic participants will maintain their performance levels. CTH predicts on the contrary that improvement may be largest for the dyslexic participants (and more generally, for the initially lowest performing children).

METHOD

Research was based in the same Warwickshire junior school. As discussed below, participants had been split in to two groups matched on the basis of age, and dyslexia 'at risk' levels. The intervention group had undertaken the DDAT exercise treatment daily at home, whereas the other group had no intervention. Following the 6 month duration of the initial study, the original controls were also offered the exercise intervention, and all accepted. They therefore formed a delayed treatment group, starting treatment 6 months later. Both groups received approximately 12 months of treatment, continuing until each participant's cerebellar/vestibular symptoms were within the normal range. The DST test was performed at 6, 12 and 24 months. The specialist tests of cerebellar/vestibular function were administered every 6 weeks until the end of each participant's programme. Because both groups had finished their remediation programmes and were no longer attending the DDAT centre, the vestibular/cerebellar tests were not repeated at 24 months. The school-administered tests of reading (NFER-Nelson) and the annual SATs measures of reading comprehension, numeracy and writing were used as additional measures of progress.

Participants

For the initial study, a request was circulated to the parents of children in 3 years of the school requesting participation in an evaluation of a novel method of literacy support. Of the 269 children, 35 were identified as potentially at risk,

using a criterion of an at risk quotient of at least 0.4 on the DST (see below for an overview of the DST). These children were then divided randomly into two groups matched for age and DST score. The groups were not formally identified at the school, and were not in any way distinguished during the school day.

The pre-test characteristics of the first intervention group and delayed intervention group were matched as follows: gender: 10m & 8f vs 9m & 8f; mean age: 9;4 vs 9;4; age range 7;11 to 10;06 vs 8;00 to 10;05; mean DST score 0.74 vs 0.72; DST range 0.4 to 1.5 vs 0.4 to 1.6.

Nine participants had a prior external diagnosis of developmental disability: mostly dyslexia (4 in the experimental group, 2 in controls), and also dyspraxia (1 in each group) and ADHD (1 control). Twelve participants (7 in the intervention group, 5 in the control group) were withdrawn for two lessons per week for small group support. This support was occurring prior to the intervention and to our knowledge these interventions remained constant over time.

Tests Administered

The dyslexia screening test

Extensive details of this test can be found in our previous paper (Reynolds *et al.*, 2003). The DST comprises 11 sub-tests in five areas (literacy skills, phonological awareness and verbal memory, motor skill and balance, and memory retrieval fluency). The sub-tests are; rapid naming, bead threading, one minute reading, postural stability, phonemic segmentation, two minute spelling, backwards digit span, nonsense passage reading, one minute writing, verbal fluency and semantic fluency. These will be abbreviated as RAN, Beads, Read, Post Stab, Seg, Spell, Memory, Nons Pass, Writing, Verbal Fl and Semantic Fl, respectively, in summary descriptions and figures.

School-administered tests

School tests included the NFER test of reading (un-timed, sentence completion, group reading Test 2) together with the three national standardized attainment tests (SATs), namely writing, comprehension and numeracy. The intervention started early in the school year (September 2001) and the school SATs testing took place in June 2001, June 2002 and June 2003. For most participants the previous year's SATs and reading data (June 2000) were also available. This allowed relative progress pre- and post-intervention to be assessed.

Procedure

Data were collected as for the previous study (Reynolds *et al.*, 2003) where a full procedural description is given. The children who had been screened as being at risk from learning difficulties were divided in to the two groups, and a battery of tests was administered at the school before treatment in line with standard practice at DDAT centres. These included the DST screening test and a range of cerebellar and vestibular tests. Based on the needs of the individual, a regime of DDAT exercises was then prescribed. Fuller details are provided in Appendix A. The exercise therapy was applied for 5–10 min twice daily every day. Examples of the 200 or so exercises used include the use of a balance stimulation board; throwing and catching of bean bags (including throwing from hand to hand with

tracking of the eye using central and peripheral vision); practice in dual tasking; spinning exercises which variously stimulate the different canals and organs of the vestibular system; stressed postural exercises which demand increasing reliance on somatosensory feedback systems; and a range of motor coordination exercises. In line with DDAT procedures, each child was tested every 6 weeks, and a new individually determined sequence of activities was prescribed for the next period. Activity sequences were designed to be cumulative, with progress based on successful completion of each stage before the next stage was attempted. The treatment is considered to be successfully completed when performance on the tests of visual and vestibular/cerebellar function rises to within the normal range. DDAT had no involvement in the collection of the post-treatment results, and tests were conducted by the University of Exeter and the school staff. It should be noted that teachers were not informed as to the groupings.

RESULTS

Data were collected at four time points: score 1 = September 2001 (pre-treatment start for initial treatment group), score 2 = March 2002 (start of treatment for delayed treatment group), score 3 = September 2002, and score 4 = September 2003. NFER and SATs test were administered in June of each year.

Analysis and Reporting Methods

The DST sub-tests each provide a 'raw score'. Using the test norms, the raw score may be converted into an age-adjusted 'at risk index' and also an age-adjusted 'decile score'. The decile score varies between 1 (the lowest performing 10% of the school population at that age) to 10 (the highest performing 10% of the population). Three SATs tests (for writing, comprehension and maths) and one set of reading test scores are also reported. The reading test is reported in terms of reading age (RA) in months. The SATs test scores have been converted into wholly numerical scores in line with the procedure used in the initial study,[†] but do not have available the normative data provided by DST and the NFER reading test. Consequently, though of fundamental importance from the school's perspective, in this scientific (rather than educational) analysis we shall treat the SATs as secondary data.

For the follow-up tests reported here, 6 of the 35 initial participants were not available for follow-up testing (with 4 moving to secondary school, one leaving the district and one leaving the programme) and so were excluded from analysis.

Performance of the Two Groups at Each Time Period

The overall decile scores on all the tests for both groups for time periods T1 and T4 are shown in Table 1. A series of analyses of covariance was undertaken on the

[†]SATs scores are normally graded in terms of levels (digits and letters). Possible grades for level 2 include (in descending order) 2a, 2b, 2c, 2, 2d. To facilitate comparison these have been coded as 2.75, 2.50, 2.25, 2.0 and 1.9, respectively. An equivalent coding was applied for levels 3 and 4.

Table 1. Means, standard deviations and between-group statistics for the two groups at T1 and T4

| | Mean (SD) T1 | Mean (SD) T4 | Effect size T4 vsT1 | Effect size gp2 vs gp 1 | Ancova |
|-----------|-----------------|-----------------|------------------------|----------------------------|--------------------------------|
| RAN | 3.60 (3.20) | 5.07 (2.76) | 0.46 | -0.17 | $F(1.24) = 0.02$ $p = 0.90$ |
| Bead | 6.00 (3.48) | 6.93 (1.79) | 0.27 | -0.26 | $F(1.24) = 0.08$ $p = 0.78$ |
| Read | 5.50 (3.11) | 7.29 (2.52) | 0.57 | -0.29 | $F(1.23) = 0.00$ $p = 0.98$ |
| Post Stab | 5.07 (2.84) | 4.80 (2.81) | -0.09 | -0.29 | $F(1.23) = 0.00$ $p = 0.98$ |
| Segment | 4.40 (2.56) | 7.07 (2.46) | 1.07 | -0.07 | $F(1.24) = 0.56$ $p = 0.46$ |
| Spell | 4.40 (2.56) | 6.07 (2.46) | 0.65 | +0.04 | $F(1.24) = 0.02$ $p = 0.89$ |
| Backspan | 3.43 (2.03) | 5.00 (1.84) | 0.78 | +0.06 | $F(1.23) = 0.58$ $p = 0.45$ |
| Nons Pass | 5.27 (2.69) | 5.73 (3.01) | 0.17 | -0.26 | $F(1.24) = 0.01$ $p = 0.93$ |
| Writing | 3.64 (1.98) | 4.00 (2.04) | 0.18 | +0.42 | $F(1.24) = 0.02$ $p = 0.90$ |
| Verbal | 5.27 (2.96) | 6.47 (3.14) | 0.40 | -0.10 | $F(1.24) = 0.08$ $p = 0.77$ |
| Fl | 4.29 (1.98) | 6.14 (2.85) | 0.94 | -0.32 | $F(1.24) = 0.32$ $p = 0.58$ |
| Semantic | 3.20 (1.74) | 3.87 (1.68) | 0.38 | -0.13 | $F(1.24) = 1.76$ $p = 0.20$ |
| Fl | 2.71 (1.44) | 2.71 (0.99) | 0.00 | | |
| Fl | 6.07 (2.81) | 6.47 (2.70) | 0.14 | | |
| Fl | 6.21 (2.15) | 6.86 (2.96) | 0.30 | | |
| Fl | 6.27 (2.99) | 5.60 (3.25) | -0.22 | | |
| Fl | 5.00 (3.31) | 5.36 (3.20) | 0.11 | | |
| Fl | 7.07 (2.87) | 8.47 (2.20) | 0.49 | | |
| Fl | 6.86 (3.16) | 8.57 (1.99) | 0.54 | | |

Data for group 1 (delayed intervention) and Group 2 (immediate intervention) are presented above each other in each cell.

data, taking performance at T1 as the covariate, performance at T4 as the dependent variable and group as the independent variable. This adjusts for any between-group differences in initial performance (see Table 1). None of the statistics approached significance, with only one F -value >1 (with $p = 0.20$) and this for semantic fluency, where both groups show large effect sizes for improvement. A referee has queried whether the study has the power to differentiate between these two groups, given a total of only 29 participants. In order to address this issue directly, we calculated the differential effect size for improvement (calculated as the difference between groups 1 and 2 in mean improvement from T1 to T4 divided by the standard deviation of the individual improvements). Effect size is independent of group size. The differential effect sizes are presented in column 5 of Table 1. Effect sizes of 0.2, 0.5 and 0.8 are considered small, medium, and large, respectively (Cohen, 1988). It may be seen that all the differential effect sizes lie below 0.42. It will be recalled that Hypothesis AH:low start proposes that the group performing most poorly is likely via statistical artifact to improve more, specifically in the literacy domain. There are four literacy tests: one minute reading, nonsense passage reading, spelling and writing. Group 2 (immediate intervention) scored less well at pre-test on all four. The effect sizes are -0.29 , $+0.42$, $+0.06$ and -0.10 , respectively. That is, one small relative advantage, one small disadvantage and two negligible

differences. Given the absence of any statistical effects, the small, random nature of the differential effect sizes, and the large effect sizes for both groups for improvement from T1 to T4, it is clear that there is no support for hypothesis AH: low-start, which may therefore be rejected. For clarity, data for the two treatment groups have therefore been pooled for the remaining analyses.

Performance of the Combined Groups at Each Time Period

Maintenance of performance, pooled across the two groups, from time T2 to T4 is shown in Table 2 (which indicates the mean decile score and standard deviation) and also displayed in Figure 1, which makes it easier to assess any trends. The population mean decile score would be 5.5. It is clear that there is no sign at T4 of regression back to performance at T1 and indeed there are signs of continuing improvement from T2 to T4 in most of the tests. *T*-tests on the change in performance from T1 to T4 indicated that there were significant improvements for vestibular/motor skills of beads and postural stability [$t = 2.37$, $p < 0.05$; $t = 4.95$, $p < 0.001$, resp.]; for verbal fluency skills of rapid automatized naming and semantic fluency [$t = 2.94$, $p < 0.01$; $t = 4.60$, $p < 0.001$, resp.] and for the verbal/cognitive skills of phonology and working memory (segmentation and backwards span) [$t = 2.95$, $p < 0.05$; $t = 2.06$, $p < 0.05$, resp.]. *T*-tests on the change from T2 to T4 indicated a significant difference only for postural stability [$t = 2.95$, $p < 0.01$]. There were in addition near-significant differences for segmentation, backwards span and semantic fluency [$t = 1.95$, $p < 0.10$; $t = 1.78$, $p < 0.10$; $t = 1.84$, $p < 0.10$, resp.]. Note that all the relative changes with time are in the direction predicted by the CTH—that is, improvement or continuing improvement.

It is evident that the DST performance improvements remained consistent over time, continuing to improve after the discontinuation of the exercise treatment for any tests for which there was an initial improvement. This is exactly as predicted by the CTH, and disconfirms hypothesis AH:Hawthorne. The continuing improvements in working memory, phonological skill and semantic fluency appear to disconfirm hypothesis AH:Transient for these sub-tests. Interestingly, though, there appears to be relatively little improvement in reading. We defer discussion of this issue until after the NFER reading data.

NFER Reading Age

The NFER reading data for the combined group are shown in Table 3. In this figure we have also displayed the data for the year (2000) before the intervention. The comparison figure of 100.0 is the mean age of the participants at the time of the initial NFER test. It may be seen that prior to the intervention the participants were behind the expected mean and decelerating in reading age (mean 7.2 months improvement in 12 months) over the period 2000–2001. There was substantial ‘acceleration’ (mean 20.0 months) in the first year of the intervention (2001–2002) for both groups, with maintenance (mean 12.0 months) after treatment ended. This led to an overall improvement of 32 months in reading age over the 24 months of the study. Data were collected for 18 participants in 2000 and 23 in subsequent years.

Table 2. Means and (standard deviations) for the DST decile scores

| | RAN | Bead | Read | Post Stab | Segment | Spell | Backspan | Nons Pass | Writing | Verbal FI | Semantic FI |
|---------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| mean T1 | 3.55 (3.05) | 5.76 (3.26) | 4.28 (2.52) | 5.07 (2.25) | 3.93 (2.33) | 4.48 (2.47) | 4.79 (2.54) | 2.97 (1.59) | 6.14 (2.47) | 5.66 (3.15) | 6.97 (2.96) |
| mean T2 | 4.52 (3.34) | 6.97 (3.33) | 4.45 (2.13) | 6.24 (1.66) | 4.79 (2.35) | 4.66 (2.69) | 5.38 (2.35) | 3.55 (1.43) | 6.24 (2.42) | 5.31 (2.99) | 7.69 (2.70) |
| mean T4 | 5.28 (3.02) | 7.10 (2.14) | 4.39 (2.44) | 7.21 (0.82) | 5.55 (2.21) | 4.93 (2.71) | 6.31 (2.95) | 3.31 (1.49) | 6.66 (2.78) | 5.48 (3.17) | 8.52 (2.06) |

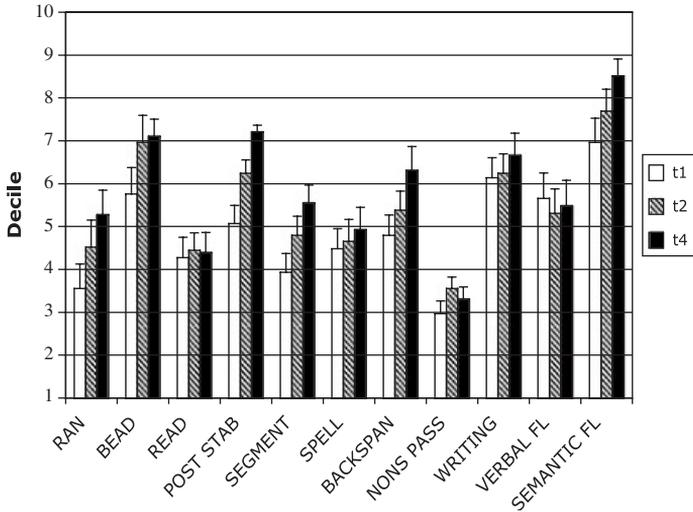


Figure 1. Changes in DST performance over time (error bars are standard errors).

Table 3. Mean scores for the two treatment groups combined compared with the national expectations, for the SAT's and NFER reading test

| | | 2000 | 2001 | 2002 | 2003 | Change 00–01 | Change 01–03 |
|-----------------------|----------|-----------------|------------------|------------------|------------------|--------------|--------------|
| NFER reading (months) | Exercise | 96.56 (17.2) | 103.74 (18.2) | 123.70 (27.1) | 135.70 (21.4) | 7.18 | 31.96 |
| | National | 100.0 | 112 | 124 | 136 | 12 | 24 |
| SAT maths (level) | Exercise | 2.56 (0.48) | 3.18 (0.64) | 3.75 (0.67) | 4.0 (0.77) | 0.62 | 0.82 |
| | National | | | | | 0.5 | 1.0 |
| SAT comp (level) | Exercise | 2.66 (0.49) | 2.91 (0.60) | 3.7 (0.67) | 4.06 (0.67) | 0.25 | 1.15 |
| | National | | | | | 0.5 | 1.0 |
| SAT writing (level) | Exercise | 2.41 (0.17) | 2.38 (0.40) | 2.92 (0.58) | 3.26 (0.52) | -0.03 | 0.88 |
| | National | | | | | 0.5 | 1.0 |

The standard deviation is given in parentheses below the mean.

Effectiveness for Participants with and without Diagnosed Developmental Disorders

We now address the fourth issue, whether the improvement was limited to non-dyslexic participants. For this purpose the combined treatment group were re-divided into two groups—those with a prior diagnosis of dyslexia or dyspraxia ($n = 8$) and those without a diagnosis of dyslexia but identified as having special educational needs by the school ($n = 21$). The DST data are shown in Figure 2.

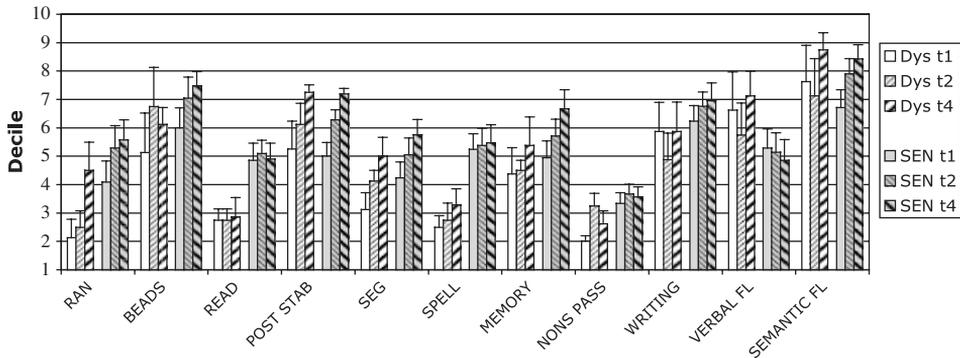


Figure 2. Changes over time of the DST sub-tests as a function of formal diagnostic category (error bars are standard errors).

It may be seen that in general the dyslexic/dyspraxic children performed worse than the SEN group on several of the DST sub-tests, but that the pattern of improvement for both groups was close to identical. Consequently this falsifies hypothesis AH:non-dyslexic.

Secondary Analyses: The SATs Scores

The above results complete the primary analyses, but it is also of interest to report briefly on further analyses which were designed for assessment of attainment rather than psychometric rigour. Given that nationally there is an expectation that SATs score should increase by 0.5 levels per year, it is possible to determine how well the intervention group kept up with this expectation. Data were collected for 11 participants in 2000 and 23 in subsequent years.

The scores on the school-administered SATs tests for the combined treatment groups are also shown in Table 3 (again indicating the changes in the year before the interventions). It may be seen that the intervention group had made good progress in mathematics before the intervention, but that their progress in the reading, comprehension and writing SATs was decelerating, given a national expectation of 1 level in 2 years. While one should not over-interpret these data, it is clear that the intervention group made roughly normal progress per year post-treatment (mean 0.48) compared with the projected mean pre-treatment of 0.28. The relatively weak improvement in mathematics is perhaps interesting, suggesting that the treatment may be more beneficial for literacy than mathematics.

DISCUSSION

Our original study assessed the 'value added' by a systematic exercise-based sequence of activities on a range of cognitive and motor skills, comparing their performance changes with a matched control group who did not undertake the exercise programme. The exercise group improved significantly more than the controls on a range of skills. Critics suggested alternative explanations for this

differential improvement. By undertaking an 18 month follow-up of the performance of the groups we have been able to evaluate how well these alternative hypotheses predict the data obtained. All four hypotheses were falsified, making predictions that failed even to capture the qualitative aspects of the data. By contrast the predictions of the CTH were largely supported, in that the exercise treatment was indeed associated with significant and lasting improvements that were maintained (and in some cases strengthened) even after the formal exercise programme terminated.

The results therefore consistently support the predictions of the CTH. Nonetheless, several provisos need to be addressed. First, and significant in the context of literacy, is the fact that there was little relative improvement on the one minute reading and the two minute spelling tests on the DST. This is perhaps puzzling in that there were strong improvements in the NFER reading test. We speculate that this difference arises because the DST literacy tests are designed to measure fluency (speed and accuracy) whereas speed is less necessary in the longer timed NFER test.

There is therefore an intriguing combination of findings, in which there are improvements in reading accuracy (NFER), phonological skill (phonemic segmentation), verbal working memory (backwards span), and language-related speed (rapid naming), but no improvement in speeded spelling, speeded reading or nonsense passage reading. In our view, this dissociation may suggest an interaction between the improvement in underlying speed, phonology and memory skills and the national literacy strategy (NLS) teaching methods used at the school. The NLS stresses the need for accuracy rather than fluency, and in particular, the phonological sounding out of words as a precursor to fluent sight reading. As established in the comprehensive studies of the National Reading Panel (NICHD, 2001) the stress on phonological decoding has an initial penalty in terms of speed. We speculate, therefore, that this dissociation between reading speed and reading fluency may be an artifact of the NLS teaching methods (see Torgesen *et al.*, 2001, pp. 53–55, for an extended discussion of why phonics instruction benefits accuracy rather than speed). It may well be that a teaching strategy that did focus on fluency and also included the reading of the occasional nonsense word might prove even more beneficial for these children.

A second issue arises from the possible specificity to dyslexia. Those participants diagnosed with dyslexia or dyspraxia made at least as good progress as those without such a diagnosis. This is of course heartening, and falsified the alternative hypothesis that the gains were attributable to the non-dyslexic participants. Nonetheless, there is something of a paradox here for the CTH. If it is indeed the cerebellar deficits that are causing the literacy problems, and the exercise programme eliminates these cerebellar problems, why do not the dyslexic group improve *more* than the non-dyslexic group? The data seem to suggest that the exercise treatment is beneficial even for those without manifest cerebellar problems.

There are three potential (and mutually compatible) resolutions of this paradox. One hypothesis is that it is possible to improve the performance of even a normally performing cerebellum by exercise treatment. This is of course entirely plausible, in that skills develop over the period from 8 to 12 years, and is supported by the continuing improvement of both groups on skills such

as postural stability. A second possibility, consistent with the fact that all participants showed physiological evidence of cerebellar/vestibular abnormality on the DDAT pre-tests, is that even non-dyslexic participants showed sub-clinical cerebellar abnormality that was insufficient to generate concern for literacy; and a third possibility is that the exercise treatment affected some non-cerebellar capability.

This latter issue was raised directly by a referee, who proposed that, despite their maintenance well after the end of the study, the long-term improvements might somehow be a Hawthorne effect or a placebo effect. The literature is somewhat confusing (a valuable Wikipedia discussion is provided by Draper, 2006), but the essence of the Hawthorne effect is that it is attributable to the knowledge that one is participating in a study and therefore puts in more effort than normal. The participants in this study did not know that they were in fact going to be re-tested 18 months later, and hence the improvement cannot be a Hawthorne Effect. Neither can it be a Pygmalion Effect (in which improvements arise because the participants' peers or teachers treat them better because they were designated as 'special' in some sense) because care was taken to ensure that participation in the study was completely independent of the school regime, and no teachers were aware of which group children were in. Nonetheless, as we discuss later, there are factors such as improved self-esteem or greater parental involvement and the like which should indeed extend beyond the completion of the intervention. These are, however, in some sense 'intrinsic by-products' of the intervention. They confuse the issue of why the improvements occurred but not whether the improvements occurred.

We have no data available on the 'affective' variables such as self-esteem and self-efficacy, though it is important to highlight the fact that the 'cerebellar cognitive affective syndrome' (Schmahmann, 2001; Schmahmann & Sherman, 1998) has been identified as a major syndrome in patients with lesions of the posterior cerebellum, and one key component of the syndrome is emotional problems.

We do have data relevant to issues at the cognitive level, and in particular whether attentional abilities might have been improved by the intervention. Attentional data were in fact collected in parallel with the DST data, following the DSM-IV requirements for diagnosis of ADHD (American Psychiatric Association, 1994), and so it is possible to check explicitly for attentional changes. Scores were based on teacher- and parent-assessments on 9 questions (such as 'easily distracted; not listen when spoken to; makes careless mistakes') leading to a score between 0 and 9, with a score of over 6 indicating ADHD. Scores were not collected at times T2 and T3, but the mean scores at T1 and T4 for the treatment groups were 4.67 and 1.93, respectively. A *t*-test between T1 and T4 indicated a significant improvement (drop) in score [$t(26) = 3.66, p = 0.001$]. Of the 27 children tested on both occasions, 12 (46%) scored 6 or more at T1, whereas only 2 (8%) scored 6 or more at T4. The effect size of the treatment was 0.97. Clearly there was a very marked (and continuing) improvement in attentional abilities following the treatment. Note that these scores are based on school-based and home-based assessments rather than 'on the spot' clinic-based tests. Limiting the analyses to the 12 participants with initial 'clinical' levels of inattention (symptom count ≥ 6) gives mean T1 and T4 scores of 8.08 and 3.08, with a corresponding effect size of 2.60.

A referee has queried whether spontaneous remission might occur with inattention symptoms. It is therefore valuable to compare these effects with published data on ADHD interventions. The National Institute of Mental Health (NIMH) Multimodal Treatment Study of Children with ADHD (The MTA Cooperative Group, 1999) investigated the effectiveness of four types of ADHD treatment. The most effective form involved monthly medication visits of a half-hour each, 35 parent training visits, a full-time 8-week summer treatment camp, 10–16 sessions of teacher consultation in behavioural principles, and 12 weeks of a half-time behaviourally trained paraprofessional aide in the child's classroom (Wells *et al.*, 2000). This led to a pre–post-treatment effect size of 1.60, smaller than obtained with the corresponding group in the present study. Caution should be exercised given the small numbers involved here and the lack of direct comparability between the two groups in terms of ADHD diagnostics, but nonetheless it is clear that the reduction in inattention symptoms is a strong finding.

Consequently, the exercise treatment was directly associated with a general improvement in attentional ability. Interestingly, this discovery appears to be directly compatible with a long-standing explanatory framework for dyslexia, the automaticity deficit framework (Nicolson & Fawcett, 1990) which holds that dyslexic children show incomplete automatization of key skills (including balance), and that therefore they need to 'consciously compensate' for these deficits in most learning situations. This need for conscious compensation would lead to a reduction in the effective mental resources available for the learning task, and would also lead to more rapid tiring. Any improvement in attentional capabilities would surely lead to long-lasting improvements in school-based learning. The automaticity deficit framework has proved valuable for other authors discussing dyslexia (Moores, Nicolson, & Fawcett, 2002; van der Leij & van Daal, 1999) and also Developmental Coordination Disorder (Visser, 2003).

An alternative but related hypothesis is that, in addition to the intended cerebellar stimulation, the exercise treatment leads to a general improvement in concentration and possibly cortical organization because of the need to undertake challenging but achievable dual tasks. Activities include manual and visual coordination (for example, throwing, visually tracking from the left visual field to the right visual field, and catching) while balancing. Given the coordinating role of the cerebellum in a range of neural systems, including the magnocellular sensorimotor system, the declarative memory system and the procedural memory system (Ullman, 2004), it would be premature to attempt to dissociate these interpretations, but nonetheless the analysis highlights the multiple neural functions that may be improved by the exercise treatment.

Further alternative hypotheses relate to a whole range of powerful 'affective' factors such as motivation, self-esteem, and self-efficacy that were the mainstay of Psychology in previous generations but have been surprisingly little considered by cognitive psychologists or educational theorists in recent years. Any such factors would of course lead to improvements in educational performance and quality of life that extended well beyond the period of the intervention. In our view, lasting changes in any of these affective factors would indeed be legitimate by-products of the intervention. Unfortunately the present study—in common with most educational interventions—did not collect the data needed to assess this affective dimension.

Finally, regardless of exactly what internal neural, cognitive and affective changes caused the improvement, it is difficult to assess the extent to which this improvement is attributable to specific cerebellar stimulation rather than say general exercise. While the effectiveness of exercise-based interventions on cognitive function was until recently considered of doubtful validity, two recent meta-analyses (Etnier *et al.*, 1997; Sibley & Etnier, 2003) of exercise-based treatments conclude that they do indeed have beneficial effects on cognitive performance, estimating mean effect sizes of 0.25 and 0.32, respectively. Of course, these effects may also be directly attributable to improvement of cerebellar function. Clearly a key requirement for the next set of studies is a design in which the 'cerebellar stimulation' treatment is contrasted with an otherwise comparable non-DDAT exercise treatment.

In conclusion, the initial study followed a 'value-added' design specifically targetting the issue of whether a commercial home-based daily exercise regime designed to stimulate the cerebellum and vestibular system was indeed likely to benefit the participants. The initial results established that there were indeed benefits on a range of cognitive and literacy skills. The follow-up study reported here has demonstrated that the majority of these gains were long-lasting, thereby falsifying the alternative hypotheses put forward by the critics of the initial study. There were three highly encouraging outcomes: the continuation of the positive effects of treatment beyond the treatment period itself; the risk group comparisons demonstrating that the effects were at least as favourable for the diagnosed dyslexic participants as for the less impaired participants; and the positive performance of these treated, learning impaired children in comparison with the national SATs expectations. To our knowledge this combination of positive factors has not been established for any intervention with this age group.

It appears clear, therefore, that the exercise treatment was associated with solid and long-lasting performance improvements. Furthermore, these improvements occurred whether or not the children had a formal diagnosis of dyslexia. This is of course particularly heartening from an applied perspective. From a theoretical perspective, however, the interpretation is less clear. It would be premature to conclude that these effects were indeed attributable solely to improved cerebellar function, given the possibility that they may be mediated by more general factors such as long-lasting improvements in attentional performance or in affective variables such as self-esteem and self-efficacy, or in social variables such as familial involvement. In short, the research reported here confirms that the exercise treatment did indeed lead to lasting benefits, but the issue of *why* requires further studies. This is an important research issue that may well lead to significant developments in our understanding of the complex inter-relationships between cognition, neuroscience, psychology and pedagogy.

APPENDIX A. FURTHER DETAILS ON THE DDAT INTERVENTION

A referee has asked for more detail on how the exercises are determined. We are grateful to Dr Roy Rutherford, the Medical Director of Dore Achievement Centres, for providing the following information. Each exercise is categorized into three types depending on the predominant sensory stimulation used throughout; those being somatosensory, visual and vestibular. Exercise types are

also categorized hierarchically depending on the measured level of sensory stimulation involved. Exercises are prescribed, through an automatized software system, in such a pattern as to stimulate two predominant senses per day so that over each 6 week session the level of sensory stimulation is equal in all three categories.

Visual stimulation is provided by means of visual exercises designed to improve rapid attention switching and foveation. Vestibular stimulation is provided by exercises including spinning round, bouncing on a gym ball 18–24 in diameter or by exercises on a round ‘wobble board’. Somatosensory stimulation is provided by exercises such as balancing on one leg with eyes closed.

The results of postural and visual neurophysiological tests (carried out every 6 or 12 weeks) and the degree of impairment on a broad range of cerebellar based neurological tests determine the start point of the programme as well as the rate at which subjects proceed through the prescriptive process. Two of the major assessment tools are the posturography equipment, which determines balance ability and adjustments under a range of conditions, including a dynamic condition in which the platform is jerked under computer control, and electronystagmography (ENG) equipment which is able to measure latency and accuracy of eye saccades and smooth pursuit. Visual performance is assessed using the ENG equipment and also by comparing balance performance with eyes closed and eyes open. Somatosensory performance is based on the ability to know the position of one’s limbs with eyes closed, and is also measured by comparing posturography conditions. Vestibular performance is assessed via measurements of nystagmus and by performance on the sway conditions in posturography.

As an example, one type of exercise in the vestibular stimulation category is ‘stressed postural stimulation’. Performance on the neurophysiological tests is used to determine the appropriate initial level of vestibular stimulation. This may start with standing for variable amounts of time on two legs with eyes closed. This progresses to a tandem stance (heel-to-toe) position with eyes open. Once this is mastered then the same stimulation is enhanced with eyes closed. This can progress through standing on one leg with eyes open then closed, to standing on a cushion (to remove somatosensory feedback and further enhance vestibular stimulation) and so on. The point at which one enters or leaves this series of stimulations will depend on the initial performance as well as the rate of serial enhancement of skilled and automatized performance at each level.

Close attention is paid to when subjects can achieve performance of these manoeuvres automatically and eventually perform them in the presence of mental distraction tasks.

References

- American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders (4th ed.)*. Washington DC: American Psychiatric Association.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Academic Press.
- Dore, W., & Rutherford, R. (2001). Closing the gap. *Paper presented at the BDA 6th international conference on dyslexia*, York, UK.
- Draper, S. (2006). *The Hawthorne effect*, from http://en.wikipedia.org/wiki/Hawthorne_Effect

- Etnier, J. L., Salazar, W., Landers, D. M., Petruzzello, S. J., Han, M., & Nowell, P. (1997). The influence of physical fitness and exercise upon cognitive functioning: A meta-analysis. *Journal of Sport & Exercise Psychology, 19*(3), 249–277.
- Fawcett, A. J., Nicolson, R. I., & Dean, P. (1996). Impaired performance of children with dyslexia on a range of cerebellar tasks. *Annals of Dyslexia, 46*, 259–283.
- Hatcher, P. J., Hulme, C., & Ellis, A. W. (1994). Ameliorating early reading failure by integrating the teaching of reading and phonological skills: The phonological linkage hypothesis. *Child Development, 65*, 41–57.
- Hook, P. E., Macaruso, P., & Jones, S. (2001). Efficacy of Fast ForWord training on facilitating acquisition of reading skills by children with reading difficulties—a longitudinal study. *Annals of Dyslexia, 51*, 75–96.
- Joseph, J., Noble, K., & Eden, G. (2001). The neurobiological basis of reading. *Journal of Learning Disabilities, 34*(6), 566–579.
- Miles, T. R. (1983). *Dyslexia: The pattern of difficulties*. Oxford: Blackwell.
- Moore, E., Nicolson, R. I., & Fawcett, A. J. (2002). Attentional deficits in dyslexia: Evidence for an automatization deficit. *European Journal of Cognitive Psychology, 15*, 321–348.
- NICHD. (2000). *Report of the national reading panel: Teaching children to read*. Washington DC: National Institute for Child Health and Human Development.
- Nicolson, R. I., & Fawcett, A. J. (1990). Automaticity: A new framework for dyslexia research? *Cognition, 35*(2), 159–182.
- Nicolson, R. I., & Fawcett, A. J. (1995). Dyslexia is more than a phonological disability. *Dyslexia: An International Journal of Research and Practice, 1*, 19–37.
- Nicolson, R. I., & Fawcett, A. J. (2005). Developmental dyslexia, learning and the cerebellum. *Journal of Neural Transmission-Supplement 69*, 19–36.
- Nicolson, R. I., & Fawcett, A. J. (2006). Do cerebellar deficits underlie phonological problems in dyslexia? *Developmental Science, 9*(3), 259–262.
- Nicolson, R. I., Fawcett, A. J., & Dean, P. (1995). Time-estimation deficits in developmental dyslexia—evidence of cerebellar involvement. *Proceedings of the Royal Society of London Series B-Biological Sciences, 259*(1354), 43–47.
- Nicolson, R. I., Fawcett, A. J., & Dean, P. (2001). Developmental dyslexia: The cerebellar deficit hypothesis. *Trends in Neurosciences, 24*(9), 508–511.
- Nicolson, R. I., & Reynolds, D. (2003a). Sound findings and appropriate statistics: Response to Snowling and Hulme. *Dyslexia, 9*, 134–135.
- Nicolson, R. I., & Reynolds, D. (2003b). Science, sense & synergy: Response to commentators. *Dyslexia, 9*, 167–176.
- Rack, J. (2003). The who, what, why and how of intervention programmes: Comments on the DDAT evaluation. *Dyslexia, 9*, 137–139.
- Ramus, F., Rosen, S., Dakin, S. C., Day, B. L., Castellote, J. M., White, S., & Frith, U. (2003). Theories of developmental dyslexia: Insights from a multiple case study of dyslexic adults. *Brain, 126*, 841–865.
- Reynolds, D., Nicolson, R. I., & Hambly, H. (2003). Evaluation of an exercise-based treatment for children with reading difficulties. *Dyslexia, 9*(1), 48–71.
- Richards, I. L., Moore, E., Witton, C., Reddy, P. A., Rippon, G., Rochelle, K. S. H., & Talcott, J. B. (2003). Science, sophistry and ‘commercial sensitivity’: Comments on ‘evaluation of an exercise-based treatment for children with reading difficulties’, by Reynolds, Nicolson and Hambly. *Dyslexia, 9*, 146–150.
- Salmelin, R., & Helenius, P. (2004). Functional neuroanatomy of impaired reading in dyslexia. *Scientific Studies of Reading, 8*(3), 257–272.
- Schmahmann, J. D. (2001). The cerebellar cognitive affective syndrome: Clinical correlations of the dysmetria of thought hypothesis. *International Review of Psychiatry, 13*(4), 313–322.
- Schmahmann, J. D., & Sherman, J. C. (1998). The cerebellar cognitive affective syndrome. *Brain, 121*, 561–579.

- Shaywitz, B. A., Shaywitz, S. E., Blachman, B. A., Pugh, K. R., Fulbright, R. K., Skudlarski, P., Mencl, W. E., Constable, R. T., Holahan, J. M., Marchione, K. E., Fletcher, J. M., Lyon, G. R., & Gore, J. C. (2004). Development of left occipitotemporal systems for skilled reading in children after a phonologically-based intervention. *Biological Psychiatry*, 55(9), 926–933.
- Shaywitz, S. E., & Shaywitz, B. A. (2005). Dyslexia (specific reading disability). *Biological Psychiatry*, 57(11), 1301–1309.
- Sibley, B. A., & Etnier, J. L. (2003). The relationship between physical activity and cognition in children: A meta-analysis. *Pediatric Exercise Science*, 15(3), 243–256.
- Singleton, C., & Stuart, M. (2003). Measurement mischief: A critique of Reynolds, Nicolson and Hambly. *Dyslexia*, 9, 151–160.
- Snowling, M. (1987). *Dyslexia: A cognitive developmental perspective*. Oxford: Blackwell.
- Snowling, M. J., & Hulme, C. (2003). A critique of claims from Reynolds, Nicolson & Hambly (2003) that DDAT is an effective treatment for children with reading difficulties—'Lies, damned lies and (inappropriate) statistics?' *Dyslexia*, 9(2), 127–133.
- Stanovich, K. E. (1988). Explaining the differences between the dyslexic and the garden-variety poor reader: The phonological-core variable-difference model. *Journal of Learning Disabilities*, 21(10), 590–604.
- Stein, J. (2003). Evaluation of an exercise-based treatment for children with reading difficulties. *Dyslexia*, 9(2), 124–126.
- Tallal, P., Merzenich, M. M., Miller, S., & Jenkins, W. (1998). Language learning impairments: Integrating basic science, technology, and remediation. *Experimental Brain Research*, 123, 210–219.
- Temple, E., Deutsch, G. K., Poldrack, R. A., Miller, S. L., Tallal, P., Merzenich, M. M., & Gabrieli, J. D. E. (2003). Neural deficits in children with dyslexia ameliorated by behavioral remediation: Evidence from functional MRI. *Proceedings of the National Academy of Sciences of the United States of America*, 100(5), 2860–2865.
- The MTA Cooperative Group. (1999). A 14-month randomized clinical trial of treatment strategies for attention-deficit/hyperactivity disorder. *Archives of General Psychiatry*, 56, 1073–1086.
- Torgesen, J. K., Alexander, A. W., Wagner, R. K., Rashotte, C. A., Voeller, K. K. S., & Conway, T. (2001). Intensive remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities*, 34(1), 33–78.
- Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, 92(1–2), 231–270.
- van der Leij, A., & van Daal, V. (1999). Automatization aspects of dyslexia: Speed limitations in word identification, sensitivity to increasing task demands, and orthographic compensation. *Journal of Learning Disabilities*, 32, 417–428.
- Vellutino, F. R., Fletcher, J. M., Snowling, M., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of Child Psychology and Psychiatry*, 45(1), 2–40.
- Visser, J. (2003). Developmental coordination disorder: A review of research on subtypes and comorbidities. *Human Movement Science*, 22(4–5), 479–493.
- Wells, K. C., Pelham, W. E., Kotkin, R. A., Hoza, B., Abikoff, H. B., Abramowitz, A., Arnold, L. E., Cantwell, D. P., Conners, C. K., Del Carmen, R., Elliott, G. R., Greenhill, L. L., Hechtman, L., Hibbs, E., Hinshaw, S. P., Jensen, P. S., March, J. S., Schiller, E., Swanson, J. M. (2000). Psychosocial treatment strategies in the MTA study: Rationale, methods, and critical issues in design and implementation. *Journal of Abnormal Child Psychology*, 28(6), 483–505.
- Wolf, M., & Bowers, P. G. (1999). The double-deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology*, 91, 415–438.
- Wolff, P. H., Michel, G. F., Ovrut, M., & Drake, C. (1990). Rate and timing precision of motor coordination in developmental dyslexia. *Developmental Psychology*, 26(3), 349–359.