

---

## ■ RESPONSE

# Science, Sense and Synergy: Response to Commentators

Roderick I. Nicolson<sup>1\*</sup>, and David Reynolds<sup>2</sup>

<sup>1</sup>*Department of Psychology, University of Sheffield, Sheffield S10 2TP, UK*

<sup>2</sup>*School of Education, University of Exeter, Heavitree Road, Exeter EX1 2LU, UK*

The commentaries on our evaluation of the DDAT exercise treatment raise several common themes and several individual themes. We rebut criticisms in terms of research design, consider the comments made, and conclude that our evidence is indeed solid. We conclude by advocating 'science, sense and synergy' as the key to making further progress. Copyright © 2003 John Wiley & Sons, Ltd.

## BACKGROUND

It has been an interesting, if dispiriting, experience surveying this set of commentaries. Perhaps the most striking aspects are, first, the adversarial approach adopted by some of the academics, and second the sheer diversity of the comments made. The paper appears to have acted as something of a Rorschach test. The authors have filtered out the parts they found relatively uninteresting—or those with which they did not find fault—and have focused on those parts in which they have special interests. The commentaries reveal as much about the authors as they do about the original paper. It would be counter-productive to produce a detailed point-by-point rebuttal of the criticisms made, in that this would merely maintain the adversarial mentality that has been, and is, so damaging to progress in the field (Nicolson, 2002). In this paper, we first summarize our original paper (as briefly as possible), then discuss the various issues raised, and then return to our primary concern as to how progress might be made.

The paper under discussion (Reynolds, Nicolson, & Hambly, 2003) was lengthy. First, we outlined issues relating to theoretical approaches to understanding dyslexia and reviewed evidence that: the cerebellum is actively involved in normal reading; it is centrally involved in acquisition and execution of motor and language skills; and there is extensive evidence of cerebellar abnormality in dyslexia. We noted that the cerebellar deficit hypothesis (see e.g. Nicolson & Fawcett, 1999) provides a natural explanation of the established

---

\*Correspondence to: Roderick I. Nicolson, Department of Psychology, University of Sheffield, Sheffield, S10 2TP, UK. E-mail: R. Nicolson@shef.ac.uk

phonological problems in dyslexia. We then moved on to intervention studies, reporting the consensus in the US and the UK that intervention for dyslexia is important, but that even the best literacy-based interventions are generally less effective and less long lasting than everyone would hope. Third, we moved on to the controversial DDAT approach, involving sophisticated clinic-based testing followed by long lasting home-based interventions using an exercise regime (including balance, eye-hand coordination, and dual task performance) designed to improve cerebellar function. The precise regime used is tailored to the child's presenting symptoms and subsequent progress. We noted the DDAT approach went beyond the cerebellar deficit theory to claim that it was possible to 'retrain' the cerebellum—the cerebellar treatment hypothesis. If valid, this would lead to generalized benefits in learning, with the particular consequence of an improvement in reading (by allowing the child to benefit more fully from the school-based teaching of reading). Finally, and this was the major thrust of the paper, we reported a school-based study of the effectiveness of the intervention.<sup>†</sup> It is important to be clear about the logic behind the study. Most well-planned interventions will have a positive effect for a variety of reasons. General effects such as improved morale, greater effort, a new start, more parental involvement and so on, are likely to lead to improved performance. In addition, each treatment will have its own specific effect—a phonological intervention will lead to better phonology, a balance intervention will lead to better balance. The key theoretical question, however, is the extent to which the intervention supports literacy. This is a particular challenge for the cerebellar treatment hypothesis, in that there is no obvious overlap between balance training and literacy, and therefore 'far transfer' to reading provided a key test of the approach. These issues formed the theoretical focus of the research. In addition, however, we were sensitive to the parental perspective. Given that the DDAT treatment is a commercial one, we wished to establish the objective value (to the child/parent) of undertaking the DDAT intervention. Finally, we were sensitive to the needs of the school. The school wished to know whether the DDAT intervention was likely to help it with its key external indicator of performance on the SATs. In order to undertake a controlled objective study it was necessary to have a control group, matched on key criteria, and experiencing the same school environment as the treatment group.

These multiple objectives led to the research design. In order to investigate specific, general and transfer effects it is necessary to administer tests of a wide range of skills. In order to investigate SATs performance one needs to use the SATs data collected by the school. Since this was done annually, whereas the study was for 6 months it was clear that these could only be of secondary importance. In order to provide a 'parent perspective' of the treatment we chose to undertake a 'value added' design that indicated what advantage a child would gain from the home-based DDAT intervention. Consequently, we compared the DDAT-treatment children with controls from the same school who undertook no additional intervention. The primary-dependent variable was the composite 'at

<sup>†</sup>In our initial submission we reported two studies, the school-based intervention and an analysis of the first 100 children going through the DDAT clinic. Following reviewing, the paper was substantially revised to include only the school-based study. The published paper was significantly modified following further refereeing.

risk' score on the 11 sub-tests of the Dyslexia Screening Test (DST) (Fawcett & Nicolson, 1996), a nationally normed test that assesses performance on a range of skills broad enough to test for specific transfer and far transfer. The two groups were matched on initial at risk score on the DST.

Several commentators have criticized the design, suggesting instead that we should have used the 'standard randomized clinical trial' intervention design, in which (i) both control and experimental groups undergo an equal time of (different) intervention (ii) control and experimental groups are matched on reading at the start (iii) (only) a range of measures of reading are used, and (iv) only children with severe reading difficulties had been used. It is important to stress that this design does not permit analysis of any of the above key issues. Introduction of a further intervention makes the issue 'is intervention A more or less effective than intervention B?' rather than 'what does a child gain from intervention A?' Furthermore, the rationale for a clinical trial is that no other treatments rather than A or B are used. This is certainly not justifiable for any educational intervention. We were well aware that a range of further interventions—at the school, and on an individual basis—were ongoing, as we documented in the original article. Second, administration only of literacy tests prevents analysis of the key issue of transfer. Matching only on reading unbalances analyses of the other skills.<sup>‡</sup> Testing only children with significant reading difficulties would have required a multi-school design. This loses a lot of experimental control because there is no reason to believe that different schools have equivalent literacy interventions. In short, it is not possible to undertake 'perfect' intervention studies. One can only design the study to investigate the key hypotheses and be clear about the limitations of the study.

In terms of the outcome of the research, we can do no better than repeat our conclusions:

In summary, the study included a sample of children selected as needing further support within a junior school, and involved home administration of the DDAT exercise regime. The progress of the DDAT intervention group was compared with that of a matched non-intervention group within the same school. Cerebellar/ vestibular signs were substantially and significantly alleviated following the DDAT 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 DDAT group in postural stability and in bead threading dexterity. There were also significant improvements in fundamental cognitive skills including phonological skill, and (one-tailed) for naming fluency and semantic fluency. Reading fluency showed a highly significant improvement for the DDAT group, and nonsense passage reading was also improved significantly. By contrast, for the control group significant age-adjusted improvement occurred only for nonsense passage reading and (one-tailed) for phonological skill. Significantly greater improvements for the DDAT group than the control group occurred for dexterity, reading, verbal fluency and semantic fluency. . . .

It should be stressed that this is only a small study, and considerably larger scale research is needed to confirm these preliminary findings and to explore the ways by which the exercise mediates the literacy improvements. Nonetheless, the results do suggest that the exercise treatment was effective, not only in its immediate target of

<sup>‡</sup>As we acknowledged in the original article, it would have been better nonetheless if the two groups had been well-matched on reading as well as DST.

improving cerebellar function but also in the more controversial role of improving cognitive skills and literacy performance.

## THE COMMENTARIES

Commentaries have been published (this issue) from Richards and colleagues, Whiteley and Pope, Singleton and Stuart, McPhillips, Rack, and Hatcher. Commentaries from Snowling and Hulme, from Peer and from Stein were published in the previous issue, along with a brief response from us (Nicolson & Reynolds, 2003) on the Snowling and Hulme paper. To avoid repetition we will group together issues across commentators and will respond only to the more salient points made.

### General Comments

#### *Experimental design*

Several commentators (Snowling & Hulme, Stein, Richards and colleagues, Singleton & Stuart) claimed that we should have used the standard intervention design. We have addressed this (Nicolson & Reynolds) and above. Our point is that we undertook a value-added design. This was necessary to investigate the 'parent-centred' view that all acknowledge is a key issue. Those wishing to undertake a 'standard design' are at liberty to do so. Indeed, having established evidence that there is a significant added value of the treatment, the logical next step (from a theoretical perspective) is to establish why there is such an effect, and how the magnitude of the effect compares with say traditional interventions.<sup>§</sup>

#### *Matching of control and intervention on reading*

Several commentators (Snowling & Hulme, Stein, Richards and colleagues, Singleton & Stuart, McPhillips) claimed that we should have explicitly matched on reading. As noted above, we matched on DST score, in that it should have given a broader match across the range of skills tested. In the event, there was a highly regrettable initial difference between the groups on reading score, as we acknowledged. We undertook the analysis of covariance suggested by Snowling and Hulme to eliminate any effect of this initial difference. This led to no change in the pattern of significant findings, thereby strengthening the results (Nicolson & Reynolds).

#### *Use of non-dyslexic children*

Several commentators (Rack, McPhillips, Stein) note that the study did not in fact test only dyslexic children. This is true. This was made clear in the original paper. For the reasons noted above, the study involved all the children in the school

<sup>§</sup>It should be noted, however, that this traditional 'Method A vs Method B' head-to-head has a remarkably poor record in terms of revealing differences (Newell, 1973). Our preference is on fact to search for the optimal method of support, which might well include a combination of Methods A and B (cf. Hatcher, Hulme, & Ellis, 1994).

who had an overall DST score of at least 0.4. It may be worth noting that the correlational analyses undertaken (Table 4) indicated that the relative improvement in reading age was not significantly correlated with original DST score (in fact the correlation is  $-0.13$ ). Clearly, if funding were available, it would be of great interest to undertake a fully controlled study with say dyslexic children, non-dyslexic poor readers, and age-matched controls.

#### *Use of SATs scores*

Several commentators (Richards and colleagues, Singleton & Stuart) query the use of SATs scores. As noted earlier, these are important for the school's purposes and provide a completely independent measure of progress. Because of the differences in timing between the end of the DDAT intervention and the SATs tests, they were never intended as a primary indicator of progress. Furthermore, both the SATs and the NFER reading tests were administered solely by the school as part of their annual testing programme. There was no possibility of retrospectively gathering further data.

#### *'No evidence'*

This appears to be something of a political catch-phrase (Snowling & Hulme, Singleton & Stuart, Rack, Richards and colleagues). It appears to be mandatory to include this phrase in one's conclusions. Presumably it means that 'the evidence presented is not sufficient'. The latter has the advantage of having a technical meaning and being open to empirical discussion. Evidence has been presented in terms of solid effect sizes, statistically significant between-group analyses (even allowing for any initial differences), and significant improvements on reading and SATs tests. The 'no evidence' position smacks of an absolutist 'blind eye' approach that is inimical to progress in the field.

### **Specific Comments**

#### *Richards and colleagues*

This is a one-sided commentary. The authors misrepresent the research hypotheses, criticize points that we did not make, and omit the major evidence in terms of effect size analyses of the DST scores. They appear to have overlooked completely the design discussion in terms of transfer effects. Most of their substantive points have been addressed above.<sup>¶</sup> However, there is one further point that we need to address.

Richards and colleagues speculate that the school's analysis of NFER reading in terms of reading ages somehow biases the results in favour of the intervention

<sup>¶</sup>The criticism on replicability (DDAT treatment is commercially sensitive and therefore it cannot be replicated by independent research) is ingenious. In fact quite the opposite holds. The DDAT procedures have been designed to be administered in a consistent way by all DDAT clinics, with decisions as to the next set of exercises for each child based upon objective diagnostic and progress information specific to each child. They are therefore objective and replicable—but only with support from DDAT. Indeed, DDAT have long-standing projects with local education authorities (and have proposed research collaboration with the Aston, York and Oxford research groups, though we understand that these offers were declined).

and argue that an analysis in terms of standard scores is the appropriate one. Given the solid effect sizes for reading age, it seemed unlikely that any different pattern would emerge with the corresponding standard scores, but in the interests of openness we carried out the proposed analysis. The standard score analyses are certainly of interest in that they do allow an age-independent measure. In the original article, we reported a comparative effect size (improvement in the year with intervention minus improvement in the year before intervention) of 1.14 for the reading ages, coupled with a significantly greater improvement in the intervention year than the previous year. The standard score data were as follows. In the year before the intervention, the intervention group's mean reading standard score declined from 98.06 to 93.63; effect size  $-0.48$ ; with the decline significant [ $t(15) = 3.00$ ,  $p < 0.01$ ]. For the intervention year, mean standard score increased from 93.63 to 99.06; effect size  $+0.62$ , with the improvement significant [ $t(15) = 2.20$ ,  $p < 0.05$ ]. The comparative effect size (subtracting that for the previous year from that for the intervention year) is therefore 1.10, and the intervention reversed the significant decline, leading instead to a significant improvement. In short, the results are if anything even stronger with the reading standard score analyses than with the reading age analyses.

#### *Singleton and Stuart*

Again, rather a tendentious tone. The discussion of screening tests is an interesting one. In fact phonemic segmentation is the DST test for phonology. However, the general point that screening tests are normally designed as 'compendium' tests (i.e. each sub-test is designed to pick up a range of potential problems) rather than targeted on a specific ability is a fair one. The criticism of the DST 1-min reading test on the grounds that it 'conflates' speed and accuracy is an interesting one. It was designed for the very purpose of combining speed and accuracy in a face-valid fashion. It is basically the number of words (taken from a list of single words increasing in difficulty) the child can read in a minute. We consider that this is actually a strength of the test, distinguishing it from standard tests of reading accuracy that (remarkably) fail to take speed into account. Anyone knows that speed is a key component of reading. The Dutch 'een-minuuten' test (Brus & Voeten, 1980), which inspired the 1-min reading test is in fact a stand-alone mainstay of Dutch reading research. We stand corrected on the SATs scoring method. It looks as though it would have been possible to get a more detailed breakdown of the SATs scores. One has to say that it is pretty difficult to envisage that this would in any way affect the pattern of results found, or the conclusions drawn.

Singleton and Stuart make a further point that 'since the intervention group were at a lower point on the overall distribution of reading ability, their opportunities for improvement were empirically and statistically better than those of the control group, regardless of treatment ... the dice were loaded in favour of reading gains by the intervention group'. If valid, this would be an important limitation. However, such a critique only arises for tests where there is a 'ceiling effect'—say the top 40% of the year all score maximum or near-maximum. The 'mirror image' problem of 'floor effects' occurs where say the bottom 40% of a cohort will score at the lowest level. Because the DST is screening

for risk, there are no floor effects. There is one DST test susceptible to ceiling effects, namely phonemic segmentation, in that there is a maximum score, and most non-dyslexic children should score at or near maximum by the age of 11 years or so. As it happens, no child in either group approached maximum on this test, and indeed the scores were pretty similar initially. In terms of the three tests on which significant between group differences were found—1 min reading, bead threading and semantic fluency—there is no possibility of ceiling effect since there is no effective scoring ceiling. In short, the issue is of relevance, but the claim made is demonstrably false.

*McPhillips*

Most of the points made are addressed above.

*Rack*

Rack provides a valuable overview for non-specialists of the issues involved. The framework in terms of how, who and what gets to the heart of the issue. We agree pretty much with most of the points made. The further analysis of covariance reported in Nicolson and Reynolds (2003) deals with Rack's main concern, namely that the effects of the intervention might derive with the initial imbalance on those tests. The only surprising (to us) aspect of the paper is the rather negative conclusions, which do not appear to follow from the rest of the paper.

*Whiteley and Pope*

We welcome the broadening of the discussion to include other exercise-based approaches, and the acknowledgement that there is a pressing need for further studies.

*Hatcher*

This paper provides an overview of recent developments of Hatcher's 'sound linkage' literacy interventions. The effects of 40 half-hours one-to-one tuition from a fully trained teacher are encouraging. The effect sizes are solid (from 0.4 to 0.8) and good progress is made in terms of reading age. These effects are in the top quartile of those reported in the literature, and given the relatively short-term (3 months) intervention are encouraging. Nonetheless, the established difficulties in terms of maintenance in improvement post-training are found. We fully support Hatcher's long-standing programme of research aimed at establishing the optimal methods of explicitly teaching literacy. The key question here is whether a parallel, complementary approach would lead to further gains and perhaps better maintenance.

*Stein (previous issue)*

Stein queries why spelling did not improve more for the DDAT group than the control group. This is an interesting fact to be explored further. There is no reason to expect reading and spelling to improve in parallel. Stein also attributes to us the idea that poor reading is caused by poor postural stability. This is an intriguing suggestion but conflates cause with correlate. We consider that both poor postural stability and poor reading are attributable to the underlying cause of poor cerebellar function. The limited evidence available from the study

suggests that the reading improvement might instead be attributable to better control of eye movements—possibly attributable either to a lack of control in the eye movement system (of which the cerebellum is a key component) or even from an inadequate vestibulo-ocular reflex, for which the cerebellum is the controller. Any interpretation in terms of eye movements would of course be consistent with the improvement found in reading but not spelling.

## SUMMARY

In summary, the commentators have noted a number of issues in relation to the study reported. Interestingly, most of the limitations raised were noted explicitly in the original article. Several commentators have focused on the issue of design. They take us to task for not using the 'standard' intervention design. We chose to use a 'value-added' design because it was simply the appropriate one for the issue of 'parental value'. We chose to use a battery of tests because they were simply the only way of addressing the issue of transfer. We chose not to employ some 'not very effective' control intervention to be done by parents because (in addition to the doubtful ethical position) we considered it diminished rather than enhanced the clarity of any findings. We made all these points explicitly in the original paper. A number of commentators were (rightly) concerned over the imbalance on initial reading age between the two groups. This was addressed by a further analysis that showed the same results obtained even when initial reading was taken into account (the commentaries predate this further analysis). There are issues relating to the SATs tests. These are not designed for psychometric purposes and were not taken immediately after the intervention. Nonetheless, they are the 'performance gold standard' for the schools and therefore inclusion of these data as a secondary measure provides a valuable indicator for education professionals.

We have undertaken all the further tests suggested by the commentators, and these have served merely to confirm the original pattern of results, further strengthening the solid evidence in terms of significantly greater learning and solid effect sizes on a range of theoretically important and educationally important indicators.

In short, we had three perspectives in mind when designing the study: theory, parents and schools. It is difficult to envisage a study that can be undertaken that simultaneously meets all these criteria (without enormous funding). In our view, the design provided a good compromise between the different perspectives. There is no doubt that the decision to match on overall DST rather than reading led to significant difficulties in interpretation, but nonetheless, this was addressed directly via the analysis of covariance. In summary, we see no reason to change our original conclusions that the exercise treatment *was* effective 'not only in its immediate target of improving cerebellar function but also in the more controversial role of improving cognitive skills and literacy performance'.

## THE WAY FORWARD

It may be worth noting that a number of studies—some small-scale, some large scale—are in progress and should cast further light on some of the issues raised

by the initial study and the commentaries. A follow-up study of progress after one further year in the Balsall Common school is now close to completion (requiring merely the scores on the summary SATs exams). The study continuation did indeed use the cross-over design advocated by Richards and colleagues (as we noted in the original paper). Furthermore, a range of studies have been undertaken in local authority schools. Finally, there is now a corpus of around 10 000 clients of the DDAT intervention. We hope to be able to report on the major findings in terms of initial characteristics and treatment progress of a substantial subset of consecutive clients to the clinic.

In our view, this is how science evolves. It does not generally advance through 'perfect' studies since there are always problems of internal and external validity, unless there is funding for massive populations in the whole range of human contexts, genetics, social and environmental. It generally advances through the accumulation of smaller-scale, necessarily partially imperfect studies that highlight one set of explanations rather than another. Confidence grows as these studies increase in number and explanatory power.

We believe that something of a paradigm shift is occurring in dyslexia research and in education. The phonological deficit hypothesis has proved a valuable descriptive and diagnostic tool and has been of considerable benefit for informing the early support of reading. Despite very extensive research there is still no agreed explanation as to why the phonological deficits arise. Despite extensive applied research, there have been few true innovations in the teaching of dyslexic children, even though there is general agreement that good support requires extensive and costly interventions. Further, an active band of dissidents has demonstrated that dyslexic children show a whole range of problems in motor skill and sensory processing that are hard to accommodate under the phonological deficit hypothesis.

It is very important that we have 'ivory tower' academics to undertake laboratory studies aimed at the eventual understanding of dyslexia. It is also a traditional academic game to criticize the research that has been done. It is only when one experiences the crushing everyday problems of dyslexic children that one realizes that academics also have a responsibility to attempt to do something to help. In our view it is better to undertake a study aimed at making progress, acknowledging that it is inevitably flawed, than to wring our hands and say 'there is no consensus on whether exercise-based remediation helps'.

We consider that the appropriate way to make progress, and to help dyslexic children, is to attempt to build rather than destroy, to listen rather than to shout down. Inevitably any single, small-scale study has limitations. Nonetheless, the study provides clear and positive evidence that the exercise treatment has (in this case) added value. Equally important, it gives us the opportunity to facilitate progress, to gather evidence, identifying the strengths and weaknesses of new approaches and attempting to synthesize the best of the new with the best of the old to combine theoretical and applied progress, moving in harness towards optimal identification and support systems based on ever-increasing scientific understanding. Science is certainly needed. Sense is however a key requisite. Above all, we need synergy between researchers with different perspectives and complementary strengths. We would particularly welcome open-minded studies of the relative merits of different interventions, both independently and in conjunction.

We believe that an optimal support platform requires (at least) three legs. Leg 1 is the explicit teaching of skills and knowledge, the traditional approach. Noone seriously suggests that this is not important. Leg 2 involves treating each child as a 'whole person', with motivation, happiness, self-image and self-belief as important considerations (Holt, 1984). The third leg, and one where specific complementary approaches may contribute, is in terms of ameliorating any impediments to learning, whether they are based on environmental, dietary or biological causes. We attempt to ignore (or kick away) these legs at our peril.

## References

- Brus, B. T., & Voeten, M. J. M. (1980). *Een-minuuttest, vorm A en B, verantwoording en handleiding*. Nijmegen: Berkhout.
- Fawcett, A. J., & Nicolson, R. I. (1996). *The Dyslexia Screening Test*. London: The Psychological Corporation.
- 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.
- Holt, J. (1984). *How children fail* (revised edition). Harmondsworth, UK: Penguin Books.
- Newell, A. (1973). You can't play 20 questions with Nature and win: Comments on the symposium. In W. Chase (Ed.), *Visual information processing*. London: Academic Press.
- Nicolson, R. I. (2002). The dyslexia ecosystem. *Dyslexia: An International Journal of Research and Practice*, *8*, 55–66.
- Nicolson, R. I., & Fawcett, A. J. (1999). Developmental dyslexia: The role of the cerebellum. *Dyslexia: An International Journal of Research and Practice*, *5*, 155–177.
- Nicolson, R. I., & Reynolds, D. E. (2003). Sound findings and appropriate statistics: Response to Snowling and Hulme. *Dyslexia: An International Journal of Research and Practice*, *9*(2), 134–135.
- Reynolds, D. E., Nicolson, R. I., & Hambly, H. (2003). Evaluation of an exercise-based treatment for children with reading difficulties. *Dyslexia: An International Journal of Research and Practice*, *9*(1), 48–71.