Further Data On The Effectiveness of the DDAT Treatment

(Paper presented to the British Dyslexia Association Conference, March 28th 2004)

David Reynolds
University of Exeter

Contact Details:
Professor David Reynolds
University of Exeter
School of Education & Lifelong Learning
Heavitree Road
Exeter
EX1 2LU
Tel: 01392 264990
Fax: 01392 264998
E-mail: David.Reynolds@exeter.ac.uk

Abstract

Work is reported on a sample of children, subjects of an earlier paper in the *Dyslexia* journal, who were clients of The Dore Achievement Centers (DDAT), who provide an individually tailored programme of exercises designed to improve cerebellar function and learning ability in children and adults with learning difficulties. These data show:

- that the children continue to reduce their dyslexia 'risk scores' over time, with the delayed treatment group performing in a similar fashion to the experimental group when they were originally treated
- improvement continues after treatment, with no 'wash-out' that often characterises such interventions
- on some achievement measures, children at high risk gain more from treatment than others
- on all achievement measures, this group of children with learning difficulties gain *more* than their age peers without such difficulties, a remarkable finding
- even after allowing for maturation, there are very satisfactory effects of treatment on cognitive speed, motor skill, balance, phonological ability, working memory and semantic fluency. Effects on reading, spelling, writing and verbal fluency were less marked. However, estimates of these latter 'literacy' skills on the national SAT's measures indicated strong improvement, at least in line with the national requirements for average annual progress, and on school administered reading tests there was an equally impressive performance.

Introduction

This is our fourth report of ongoing work to evaluate the treatment regime of DDAT. Our first two studies were initial exploratory work that looked at whether those who were self reporting as having done more of the exercise regime made more progress when self rating the quality of their lives, which was indeed the case. The second study took a small sample of about 100 of the first children who had experienced the treatment and looked at whether their 'at risk' scores on the industry standard dyslexia test (DST) declined over time, by comparison with what would have been expected from those without treatment, taken from a national study on which the DST was standardised and normed. This was also the case.

Our third study was a considerable scientific advance on this earlier work, and was published in the peer reviewed journal *Dyslexia*. It reported on an experiment in which the DDAT treatment was given to a group of 'experimental' children with a similar 'control' group not receiving the treatment until six months later, a variant on the 'crossover' design so popular in clinical trials. The experimental group improved their DDAT 'risk' score much more than the control group, and their superiority also extended to performance on the Comprehension and Writing elements of the national SATs, where their performance improved much more rapidly during treatment than before it.

This study was based, however, on small sample sizes (35 children in total for the two groups) and only involved reporting of the results of the experimental group, not of the control group when they became a ‘delayed treatment’ group too, from Spring 2002. It was also unclear from this study whether the observed improvement would hold over time, or 'wash out' in the way that other intervention programmes for dyslexia and related problems have demonstrated. The treatment samples were also not compared to other children, so we were unsure of the extent to which simple maturational change may be generating the improved scores.

In this report we look at:

- Whether the ‘delayed treatment’ group showed positive treatment effects, like the first experimental group.

- Whether any gains due to treatment were maintained after treatment ceased.

- Whether those with initially a higher degree of 'risk' on the DST gained more or less from treatment.

- Whether gains were in excess of those seen in non-learning impaired children at the same school.

We use data on:

- The 'Balsall Common' children, who were given the DST tests in Summer 2001, Spring 2002, Autumn 2002 and Summer 2003, and who sat national SATs, and a special school administered NFER reading test, in Summer 2000, 2001, 2002 and 2003

We should note that the DST test that we have utilised extensively in this research is a reputable test of demonstrable reliability and validity that can be administered at intervals of six months or more, as we have utilised it here. All the sub scales have been used as 'stand alone' tests of various of the skill / development areas that dyslexic children can experience difficulties with.

We should also note that the teachers in this study were 'blind' as to which children were in the study, and obviously as to which children were in each of the two groups.

The Balsall Common Study

Data here are firstly based upon DDAT overall 'at risk' scores and further on the sub scales that together give the overall score. **Table One** shows that the two experimental or treatment groups improve from an overall risk score of 0.65 at start of treatment to one of 0.31 in Autumn 2002 (when the first group treated are ceasing treatment). Crucially, the first experimental group continue to improve between Autumn 2002 and 2003 even though they are no longer on treatment, suggesting no treatment 'wash out'. Both groups move as a whole from a 'moderate risk' level to very close to the 'no risk' cut off at 0.20 (We should make clear that the first experimental group started in Autumn Term 2001 and stopped at the end of the Spring Term 2002. The second experimental group started at the beginning of the Summer Term 2002 and finished at the end of the Autumn Term 2002).

These improvements validate our original suggestions of a significant positive treatment effect by the second 'delayed treatment' group improving when treated and the first experimental group, remarkably, continuing to improve when treatment ceases, an effect we have never seen before in the literature on dyslexia treatments.

Performance of the two groups on the eleven sub scales of the DST tests are also included as part of **Table One**.

We have also analysed the Balsall Common data across the two groups to see if children more 'at risk' initially tend to gain more or less from treatment. **Table Two** shows that children with more severe symptoms of dyslexia, as diagnosed by the DST, gain significantly more from treatment. Scores are given for the overall DST, and for each of the sub scales. The 'low risk' group improve from a DST score of 0.41 to one of 0.13, suggesting their mild dyslexia has been eradicated. The 'high risk' group move from a score of 0.92, close to the 'severe risk' level of 1.0, to a score of 0.34.

There is a further set of data that it is important to look at, which is the data collected routinely by the school itself in its own internal testing programme, and through the national testing programme of SATs. These data are important in that they indicate whether the intervention is generating improvement not merely in the underlying skills that influence literacy but also whether there is transfer to the process of reading itself. These data are also important in that they are collected by others than the research team, generating an external check on the validity of any findings generated by data collected especially for the purpose of the project evaluation. **Table Three** shows the data for the school administered reading test, with mean reading age for the sample measured in months at four different time points (note that sample sizes here are below those reported elsewhere in this paper and in our earlier report on the Balsall Common data, since we are here only using data from two out of the three years of pupils who formed the original sample of 35, the pupils from the eldest group of children having gone to secondary school and passed out of the school's testing regime. Five children were also deleted from the sample as they did not complete the exercise regime).

These data show a very similar positive pattern to that for the DST -- a very rapid rise in reading ages for the first and second experimental groups as they are treated in 2001/2 by comparison with both groups gains in the previous academic year, and not just the maintenance of gains by these experimental groups in the year after treatment, but further gains considerably in excess of those that would be anticipated due to maturation and considerably in excess of those being exhibited before treatment. For the whole sample, by the academic year 2002/3 they have been elevated to a level of increase in reading performance roughly to be expected of children of their age, a rate of increase double that which they were showing before treatment.

For the SATs results, the picture is more mixed. On Mathematics, the level of increase stays similar on treatment to that exhibited before it, and then slows as there appears to be washout from the experimental group. On writing, the rate of increase in years 2001/2 and 2002/3 greatly exceeds a decline in scores over time prior to treatment. On comprehension, the substantial increase when treated for both experimental and 'delayed treatment' groups still continues after treatment, although at a lower level than during treatment. For both these skill areas, the rate of increase is higher both during and after treatment than before treatment. We look later at the extent to which this variation by subject area was prevalent in the rest of the school population, but we should note that on the SAT's, the mean improvement for the treated groups are 0.86, 1.19 and 0.91 for mathematics, comprehension and writing, very close to the improvement level of 0.5 of a national curriculum level per pupil per year (1.0 of a level over two years) that the national assessment system is based upon. For reading age, the improvement is almost 32 months in 24 months.

Analysis, as in **Table Five**, of these internal school data and external SATs data in terms of the performance of pupils with different levels of risk scores shows a similar set of results to that for the whole sample (note that ANOVAs are only carried out for the last three years of the time period because sample sizes for 2000 are very low). On the school's own reading comprehension measure, those children across the two groups, experimental and control, with higher risk scores progressed faster than the others in the treatment year 2001-2 and afterwards in 2002/3, with the higher risk children finishing only one and a half months behind their lower risk peers.

On the SATs, the Mathematics scores increased less for those at high risk than for the other group, a finding paralleled in the Writing SATs. On Comprehension, though, the increase was greater for the group at high risk than for that at low risk.

We should add that a proportion of children across the two groups – perhaps a quarter – were already receiving, when the intervention began, various kinds of support for their learning problems, ranging from the intervention of an educational psychologist to the efforts of mainstream teachers in classrooms. To our knowledge, these interventions remained constant over time and the improvement in children on the DDAT treatment is a result of the treatment itself, not these other interventions.

We should also note that the poorer levels of impact upon mathematics achievement may reflect upon the subject specificity of maths in that progress in this hierarchical subject requires foundational skills in order to make progress. By contrast to literacy, where if early skills are not well developed there is still a chance of remediating later, mathematics may need the development of prior, basic skills that are necessary for development of more advanced maths skills.

It is important to note that we have conducted two final analyses of the Balsall Common data that increase the sensitivity of our analyses. In the first one, we have looked not just at the gain over time in the two groups of children who have been treated, which is impressive, but at the gain over time of the children in the school who were not being treated because they had no learning disability and who were the 'normal' non-dyslexic school population. This is to take account of the possibility that the substantial improvement in the two groups of children who were remediated in the programme (the 'experimental' and 'control' groups) may have been due to simple maturation that is typical for children attending the research school, Balsall Common Primary. It is important to add that we have done this further analysis because of the possibility of the treated children's substantial gains being due to being in a high quality school, with a pupil balance of children from middle class backgrounds, which may have been implicated in their progress.

Table Six shows data confirming that the substantial improvement in reading ages and SAT's scores is in all cases greater for the treated children than their peers on the raw data, a remarkable result given the characteristics of the two groups. If allowance is made for the initial start point of the various groups of 'treated' children and for their peer group, then the percentage increase is considerably greater for the treated children than their age peers. In the case of reading for example, the percentage improvement is approximately

30% of the start score for the treated children, and approximately 15% for their peer group of children from the same years at school, a doubled progress rate.

Our second final analysis looks at the extent to which the substantial improvements in the DST 'risk' scores of the treated children simply reflect their maturation and ageing. We noted in our earlier discussion of the data in **Table One** that there is a substantial reduction in the various sub scale scores. **Table Seven** shows the effect sizes of this reduction on raw scores on the DST, calculated as being the amount of improvement divided by the averaged standard deviation. It is generally agreed that an effect size of 0.5 can be considered 'moderate' and one of 0.8 or more can be considered 'large'. Whilst the 'literacy based' sub scales show overall lower effect sizes that the others, there is a large effect size overall across all the sub scales.

Table Eight proceeds to show the effect sizes of change in the two groups, utilising decile scores which are a more appropriate way of assessing improvement over time and which take account of maturation.

As might be expected, there are still very large effect sizes for postural stability and bead threading, relating to the effect of the treatment upon fine and gross motor skills. There are still 'moderate' or 'close to large' effects upon aspects of general cognitive performance such as segmentation skills, rapid naming, backward memory span and semantic fluency, which indicates phonological, semantic and working memory improvements. These skills could be regarded as 'fundamental' for literacy skills.

The literacy based sub tests show smaller effect sizes, however, in contrast to the large effect sizes for the school administered reading tests. It may be that the one minute reading test, the two minute spelling test and the one minute writing test necessitate rapid reaction times that dyslexic children find hard to achieve, which are less necessary in the longer timed NFER test. It may also be particularly that the NFER test involves assessment of fluency, comprehension and word decoding skills, in contrast to the one minute reading tests concentration upon fluency and decoding.

Conclusions

In this study, we have significantly added to the data that we originally possessed about the possible effectiveness of the DDAT intervention, in terms of more detailed analysis of our initial sample.

In general, our results are supportive of the treatment. The group of children in our original study who were yet to get the treatment, get it and improve significantly on their DST risk scores, reading scores and the reading element of their SATs. On the school reading test and the comprehension SAT's, those with the lowest potential -- those with the highest risk scores or lowest achievement scores -- gain more from the intervention than those with an initially more favourable prognosis. This is not seen in other treatments for dyslexia, where normally those with the most severe risk make least progress.

We also looked at the extent to which the treated children perform on achievement measures by comparison to their age peers of non-learning disabled children. The treated children, behind on every measure of achievement we have utilised, make more progress than their peers. We have never seen this in any other dyslexic interventions either and, indeed, sadly we have never seen a study that even attempts to compare dyslexic treated groups with normally developing subjects. There are also strong suggestions of the continuation of the positive effects of treatment beyond the treatment period itself, without 'washing out', in reading, comprehension and writing. We have never seen this in other existing treatments either.

These major 'positives' sit against smaller effect sizes from the age adjusted DST 'literacy' measures, although whether this reflects the mechanics of the testing procedure is unclear. On the other DST sub scales, the effects are positive.

Weighing all the data together, it is difficult to avoid the conclusion that the DDAT treatment is achieving permanent change in the lives of the children who are on it.

Critics have attempted to argue that any positive effects of treatment are due to an expectancy effect that operates to improve performance because the children are being given attention of a general kind, rather than because of the precise characteristics of the DDAT treatment. This is a most unlikely explanation, since very many attempts to improve the performance of dyslexic children have been made, all of which would have had general effects, yet all of which have been disappointing. Also, if the results were due to a generalised 'expectancy' or 'treatment' effect, then there should be no variation between the specific subjects tested, as is the case shown in these data by the difference between skill areas. Simply, the DDAT intervention does not work because it is an intervention, or this would be shown by all interventions. It works because of its characteristics.

Note:

There are other datasets that we are at present working on, including the results of patients 'self-reporting' on the effects of treatment and data from the progress of large samples of patients through the treatment, where we have been hindered by changes in the administration of the DST test from administration of the whole DST to administration only of a number of sub-scales on the majority of the population. However, the recent reassertion of the practice of administration of the whole DST as well as professional assessments of ADHD and DCD (Developmental co-ordination disorder), using a highly developed computerised data collection system which is now fully operational, will allow assessment and analysis of many thousands of records.

<u>Table One: Balsall Common DST Results</u>

(Note that in this and successive descriptions of the overall DST scores, low scores indicate lower risk)

Mean DST Scores

Group	Score 1	Score 2	Score 3	Score 4
Control	.61 (n=15, *SD* =.30)	.41 (n=15, *SD*=.43)	.30 (n=14, *SD*=.30)	.22 (n=15, *SD*=.21)
Experimental	.71 (n=14, *SD* =.33)	.37 (n=14, *SD*=.25)	.31 (n=14, *SD*=.16)	.25 (n=13, *SD*=.16)
Total	.65 (n=29, *SD*=.32)	.39 (n=29, *SD*=.35)	.31 (n=28, *SD*=.24)	.24 (n=28, *SD*=.19)

A repeated measures ANOVA showed a significant main effect of Score, F (3, 75) = 25.17, $p < .001$, no significant main effect of Group, F (1, 25) = 0.03, $p = .868$, and no significant interaction of Score and Group, F (3, 75) = 0.84, $p = .477$.

Mean Rapid Naming Scores
(Low scores indicate improved outcomes)

Group	Score 1	Score 2	Score 3	Score 4
Control	46.60 (n=15, *SD* =11.90)	43.27 (n=15, *SD*=13.88)	38.79 (n=14, *SD*=8.05)	34.47 (n=15, *SD*=6.55)
Experimental	46.36 (n=14, *SD* =11.03)	41.93 (n=14, *SD*=11.23)	38.21 (n=14, *SD*=8.62)	35.36 (n=14, *SD*=9.37)
Total	46.48 (n=29, *SD*=11.28)	42.62 (n=29, *SD*=12.46)	38.50 (n=28, *SD*=8.19)	34.89 (n=29, *SD*=7.90)

A repeated measures ANOVA showed a significant main effect of Score, F (3, 78) = 14.49, $p < .001$, no significant main effect of Group, F (1, 26) < 0.001, $p = .996$, and no significant interaction of Score and Group, F (3, 78) = 0.19, $p = .906$.

Mean Bead Threading Scores

Group	Score 1	Score 2	Score 3	Score 4
Control	7.47 (n=15, *SD* =1.96)	7.40 (n=15, *SD*=1.35)	8.71 (n=14, *SD*=1.33)	9.13 (n=15, *SD*=1.25)
Experimental	7.36 (n=14, *SD* =1.39)	8.79 (n=14, *SD*=1.25)	8.79 (n=14, *SD*=1.25)	9.21 (n=14, *SD*=1.48)
Total	7.41 (n=29, *SD*=1.68)	8.07 (n=29, *SD*=1.46)	8.75 (n=28, *SD*=1.27)	9.17 (n=29, *SD*=1.34)

A repeated measures ANOVA, with adjustment for the Greenhouse-Geisser effect, showed a significant main effect of Score, F (2.45, 63.65) = 15.02, $p < .001$, no significant main effect of Group, F (0.82, 21.22) = 0.71, $p = .409$, and a significant interaction of Score and Group, F (2.45, 63.65) = 3.31, $p < .05$.

Mean One Minute Reading Scores

Group	Score 1	Score 2	Score 3	Score 4
Control	38.87 (n=15, *SD* =21.77)	43.67 (n=15, *SD*=18.79)	50.29 (n=14, *SD*=20.19)	57.80 (n=15, *SD*=21.76)
Experimental	23.07 (n=14, *SD* =14.80)	32.64 (n=14, *SD*=13.06)	37.93 (n=14, *SD*=16.18)	47.15 (n=14, *SD*=17.78)
Total	31.24 (n=29, *SD*=20.08)	38.34 (n=29, *SD*=16.95)	44.11 (n=28, *SD*=19.03)	52.86 (n=29, *SD*=20.38)

A repeated measures ANOVA showed a significant main effect of Score, F (3, 75) = 80.68, $p < .001$, no significant main effect of Group, F (1, 25) = 3.17, $p = .087$, and no significant interaction of Score and Group, F (3, 75) = 1.52, $p = .215$.

Mean Postural Stability Scores
(Low scores indicate improved outcomes)

Group	Score 1	Score 2	Score 3	Score 4
Control	6.27 (n=15, *SD* =6.04)	3.40 (n=15, *SD*=4.00)	0.50 (n=14, *SD*=0.85)	0.20 (n=15, *SD*=0.77)
Experimental	5.64 (n=14, *SD* =3.77)	1.93 (n=14, *SD*=2.13)	0.57 (n=14, *SD*=1.16)	0.93 (n=14, *SD*=1.64)
Total	5.97 (n=29, *SD*=5.00)	2.69 (n=29, *SD*=3.26)	0.54 (n=28, *SD*=1.00)	0.55 (n=29, *SD*=1.30)

A repeated measures ANOVA, with adjustment for the Greenhouse-Geisser effect, showed a significant main effect of Score, F (1.82, 47.42) = 21.77, $p < .001$, no significant main effect of Group, F (0.61,15.81) = 0.44, $p = .515$, and no significant interaction of Score and Group, F (1.82, 47.42) = 0.86, $p = .419$.

Mean Phonemic Segmentation Scores

Group	Score 1	Score 2	Score 3	Score 4
Control	8.53 (n=15, *SD* =1.81)	9.27 (n=15, *SD*=2.15)	10.21 (n=14, *SD*=1.67)	10.53 (n=15, *SD*=1.55)
Experimental	7.43 (n=14, *SD* =1.87)	8.86 (n=14, *SD*=2.03)	10.43 (n=14, *SD*=1.28)	9.93 (n=14, *SD*=1.07)
Total	8.00 (n=29, *SD*=1.89)	9.07 (n=29, *SD*=2.07)	10.32 (n=28, *SD*=1.47)	10.24 (n=29, *SD*=1.35)

A repeated measures ANOVA showed a significant main effect of Score, F (3, 78) = 30.01, $p < .001$, no significant main effect of Group, F (1, 26) = 0.52, $p = .478$, and no significant interaction of Score and Group, F (3, 78) = 1.51, $p = .218$.

Mean Two Minute Spelling Scores

Group	Score 1	Score 2	Score 3	Score 4
Control	15.47 (n=15, *SD* =4.79)	17.73 (n=15, *SD*=5.42)	19.50 (n=14, *SD*=6.16)	21.13 (n=15, *SD*=5.28)
Experimental	12.64 (n=14, *SD* =3.54)	14.00 (n=14, *SD*=4.77)	18.71 (n=14, *SD*=7.68)	18.62 (n=14, *SD*=3.59)
Total	14.10 (n=29, *SD*=4.40)	15.93 (n=29, *SD*=5.37)	19.11 (n=28, *SD*=6.84)	20.50 (n=29, *SD*=4.83)

A repeated measures ANOVA, with adjustment for the Greenhouse-Geisser effect, showed a significant main effect of Score, F (1.52, 38.03) = 17.14, $p < .001$, no significant main effect of Group, F (0.51, 12.68) = 2.11, $p = .159$, and no significant interaction of Score and Group, F (1.52, 38.03) = 0.77, $p = .439$.

Mean Backwards Span Scores

Group	Score 1	Score 2	Score 3	Score 4
Control	3.67 (n=15, *SD* =1.76)	4.53 (n=15, *SD*=1.68)	4.93 (n=14, *SD*=1.90)	5.33 (n=15, *SD*=1.99)
Experimental	3.07 (n=14, *SD* =1.00)	3.50 (n=14, *SD*=0.76)	4.57 (n=14, *SD*=1.16)	4.79 (n=14, *SD*=1.12)
Total	3.38 (n=29, *SD*=1.45)	4.03 (n=29, *SD*=1.40)	4.75 (n=28, *SD*=1.55)	5.07 (n=29, *SD*=1.62)

A repeated measures ANOVA showed a significant main effect of Score, F (3, 78) = 11.03, $p < .001$, no significant main effect of Group, F (1, 26) = 2.50, $p = .126$, and no significant interaction of Score and Group, F (3, 78) = 0.44, $p = .726$.

Mean Nonsense Passage Scores

Group	Score 1	Score 2	Score 3	Score 4
Control	48.20 (n=15, *SD* =13.06)	56.47 (n=15, *SD*=10.23)	57.07 (n=14, *SD*=11.26)	62.87 (n=15, *SD*=10.27)
Experimental	44.50 (n=14, *SD* =9.73)	53.71 (n=14, *SD*=7.46)	53.14 (n=14, *SD*=9.12)	57.32 (n=14, *SD*=7.38)
Total	46.41 (n=29, *SD*=11.52)	55.14 (n=29, *SD*=8.95)	55.11 (n=28, *SD*=10.25)	60.19 (n=29, *SD*=9.27)

A repeated measures ANOVA showed a significant main effect of Score, F (3, 78) = 36.64, $p < .001$, no significant main effect of Group, F (1, 26) = 1.30, $p = .265$, and no significant interaction of Score and Group, F (3, 78) = 0.57, $p = .635$.

Mean One Minute Writing Scores

Group	Score 1	Score 2	Score 3	Score 4
Control	15.40 (n=15, *SD* =6.24)	16.87 (n=15, *SD*=4.41)	18.43 (n=14, *SD*=5.94)	20.93 (n=15, *SD*=4.71)
Experimental	15.86 (n=14, *SD* =6.10)	17.07 (n=14, *SD*=5.55)	19.21 (n=14, *SD*=7.05)	21.57 (n=14, *SD*=5.52)
Total	15.62 (n=29, *SD*=6.07)	16.97 (n=29, *SD*=4.90)	18.82 (n=28, *SD*=6.41)	21.24 (n=29, *SD*=5.03)

A repeated measures ANOVA showed a significant main effect of Score, F (3, 78) = 13.45, $p < .001$, no significant main effect of Group, F (1, 26) = 0.14, $p = .710$, and no significant interaction of Score and Group, F (3, 78) = 0.06, $p = .981$.

Mean Verbal Fluency Scores

Group	Score 1	Score 2	Score 3	Score 4
Control	11.40 (n=15, *SD* =4.36)	10.87 (n=15, *SD*=4.31)	12.79 (n=14, *SD*=3.24)	13.47 (n=15, *SD*=5.03)
Experimental	9.07 (n=14, *SD* =4.21)	10.36 (n=14, *SD*=3.77)	12.43 (n=14, *SD*=2.95)	12.43 (n=14, *SD*=4.54)
Total	10.28 (n=29, *SD*=4.37)	10.62 (n=29, *SD*=3.99)	12.61 (n=28, *SD*=3.05)	12.97 (n=29, *SD*=4.74)

A repeated measures ANOVA showed a significant main effect of Score, F (3, 78) = 7.61, $p < .001$, no significant main effect of Group, F (1, 26) = 0.79, $p = .382$, and no significant interaction of Score and Group, F (3, 78) = 0.86, $p = .467$.

Mean Semantic Fluency Scores

Group	Score 1	Score 2	Score 3	Score 4
Control	14.87 (n=15, *SD* =4.07)	15.00 (n=15, *SD*=3.96)	19.79 (n=14, *SD*=8.13)	19.07 (n=15, *SD*=4.20)
Experimental	15.07 (n=14, *SD* =5.33)	19.00 (n=14, *SD*=5.32)	17.57 (n=14, *SD*=4.03)	19.29 (n=14, *SD*=4.43)
Total	14.97 (n=29, *SD*=4.63)	16.93 (n=29, *SD*=5.01)	18.68 (n=28, *SD*=6.39)	19.17 (n=29, *SD*=4.23)

A repeated measures ANOVA, with adjustment for the Greenhouse-Geisser effect, showed a significant main effect of Score, F (2.22, 57.80) = 6.59, $p < .005$, no significant main effect of Group, F (0.74, 19.27) = 0.25, $p = .620$, and no significant interaction of Score and Group, F (2.22, 57.80) = 2.78, $p = .065$.

Table Two: Balsall Common DST Results By Risk Group

DST Subscale Scores

Risk Group 1: DST1 Score of 0 to 0.5, N = 14
Risk Group 2: DST1 Score of 0.6 to 1.5, N = 14

Mean DST Scores

Risk Group	Score 1	*SD* 1	Score 2	*SD* 2	Score 3	*SD* 3	Score 4	*SD* 4
1	0.41	0.08	0.16	0.17	0.21	0.17	0.13	0.09
2	0.92	0.25	0.64	0.32	0.41	0.25	0.34	0.20

A repeated measures ANOVA showed a significant main effect of Score, $F(3, 75) = 29.38$, $p < .001$, a significant main effect of Group, $F(1, 25) = 53.03$, $p < .001$, and a significant interaction of Score and Group, $F(3, 75) = 6.25$, $p < .005$.

Mean Rapid Naming Scores

Risk Group	Score 1	*SD* 1	Score 2	*SD* 2	Score 3	*SD* 3	Score 4	*SD* 4
1	39.14	9.91	37.00	11.60	36.14	7.83	32.00	6.04
2	52.64	7.49	48.07	11.55	40.86	8.12	37.79	8.94

A repeated measures ANOVA showed a significant main effect of Score, $F(3, 78) = 16.10$, $p < .001$, a significant main effect of Group, $F(1, 26) = 10.07$, $p < .005$, and a significant interaction of Score and Group, $F(3, 78) = 3.11$, $p < .05$.

Mean Bead Threading Scores

Risk Group	Score 1	*SD* 1	Score 2	*SD* 2	Score 3	*SD* 3	Score 4	*SD* 4
1	7.29	2.13	7.79	1.58	8.86	1.35	9.07	1.59
2	7.57	1.22	8.36	1.39	8.64	1.22	9.29	1.14

A repeated measures ANOVA showed a significant main effect of Score, $F(3, 78) = 13.64$, $p < .001$, no significant main effect of Group, $F(1, 26) = .25$, $p = .622$, and no significant interaction of Score and Group, $F(3, 78) = .61$, $p = .612$.

Mean One Minute Reading Scores

Risk Group	Score 1	*SD* 1	Score 2	*SD* 2	Score 3	*SD* 3	Score 4	*SD* 4
1	37.93	22.12	44.00	18.63	48.79	20.43	61.85	21.35
2	24.29	16.67	32.14	13.82	39.43	16.95	44.64	16.97

A repeated measures ANOVA showed a significant main effect of Score, $F(3, 75) = 82.05$, $p < .001$, no significant main effect of Group, $F(1, 25) = 3.90$, $p = .06$, and no significant interaction of Score and Group, $F(3, 75) = 2.07$, $p = .111$.

Mean Postural Stability Scores

Risk Group	Score 1	*SD* 1	Score 2	*SD* 2	Score 3	*SD* 3	Score 4	*SD* 4
1	6.14	4.29	3.14	3.78	0.50	0.85	0.36	1.37
2	6.07	5.84	2.36	2.84	0.57	1.16	0.79	1.31

A repeated measures ANOVA, with adjustment for the Greenhouse-Geisser effect, showed a significant main effect of Score, $F(1.86, 48.36) = 21.23$, $p < .001$, no significant main effect of Group, $F(0.62, 16.12) = 0.01$, $p = .897$, and no significant interaction of Score and Group, $F(1.86, 48.36) = 0.20$, $p = .804$.

Mean Phonemic Segmentation Scores

Risk Group	Score 1	*SD* 1	Score 2	*SD* 2	Score 3	*SD* 3	Score 4	*SD* 4
1	8.07	1.90	9.57	1.83	10.29	1.64	10.21	1.48
2	7.79	1.93	8.43	2.21	10.36	1.34	10.14	1.23

A repeated measures ANOVA showed a significant main effect of Score, F(3, 78) = 30.30, $p < .001$, no significant main effect of Group, F(1, 26) = 0.43, $p = .519$, and no significant interaction of Score and Group, F(3, 78) = 1.78, $p = .157$.

Mean Two Minute Spelling Scores

Risk Group	Score 1	*SD* 1	Score 2	*SD* 2	Score 3	*SD* 3	Score 4	*SD* 4
1	15.00	3.84	16.79	4.81	18.71	5.50	21.62	18.86
2	12.64	4.36	14.57	5.69	19.50	8.17	18.86	3.37

A repeated measures ANOVA with adjustments for the Greenhouse-Geisser effect, showed a significant main effect of Score, F(1.54, 38.40) = 17.21, $p < .001$, no significant main effect of Group, F(0.512, 12.80) = 1.09, $p = .306$, and no significant interaction of Score and Group, F(1.54, 38.40) = 1.18, $p = .308$.

Mean Backwards Span Scores

Risk Group	Score 1	*SD* 1	Score 2	*SD* 2	Score 3	*SD* 3	Score 4	*SD* 4
1	3.64	1.82	4.64	1.69	4.93	1.64	5.79	1.85
2	3.14	1.03	3.43	0.76	4.57	1.50	4.36	1.08

A repeated measures ANOVA showed a significant main effect of Score, F(3, 78) = 11.43, $p < .001$, a significant main effect of Group, F(1, 26) = 4.73, $p < .05$, and no significant interaction of Score and Group, F(3, 78) = 1.40, $p = .248$.

Mean Nonsense Passage Scores

Risk Group	Score 1	*SD* 1	Score 2	*SD* 2	Score 3	*SD* 3	Score 4	*SD* 4
1	51.00	9.27	56.71	7.27	56.14	10.71	63.00	9.12
2	41.43	12.18	53.14	10.49	54.07	10.06	57.54	9.25

A repeated measures ANOVA showed a significant main effect of Score, $F(3, 78) = 40.15$, $p < .001$, no significant main effect of Group, $F(1, 26) = 2.34$, $p = .138$, and a significant interaction of Score and Group, $F(3, 78) = 3.11$, $p < .05$.

Mean One Minute Writing Scores

Risk Group	Score 1	*SD* 1	Score 2	*SD* 2	Score 3	*SD* 3	Score 4	*SD* 4
1	17.21	6.66	19.64	4.57	22.43	6.00	23.29	4.08
2	13.64	5.08	14.29	3.93	15.21	4.63	18.93	5.15

A repeated measures ANOVA showed a significant main effect of Score, $F(3, 78) = 14.18$, $p < .001$, a significant main effect of Group, $F(1, 26) = 10.82$, $p < .005$, and no significant interaction of Score and Group, $F(3, 78) = 1.46$, $p = .231$.

Mean Verbal Fluency Scores

Risk Group	Score 1	*SD* 1	Score 2	*SD* 2	Score 3	*SD* 3	Score 4	*SD* 4
1	11.29	4.45	11.86	2.68	12.71	2.61	12.93	4.83
2	9.29	4.39	9.43	4.89	12.50	3.52	13.21	4.93

A repeated measures ANOVA showed a significant main effect of Score, $F(3, 78) = 7.88$, $p < .001$, no significant main effect of Group, $F(1, 26) = 0.69$, $p = .413$, and no significant interaction of Score and Group, $F(3, 78) = 1.79$, $p = .156$.

Mean Semantic Fluency Scores

Risk Group	Score 1	*SD* 1	Score 2	*SD* 2	Score 3	*SD* 3	Score 4	*SD* 4
1	15.29	4.43	19.00	4.80	18.07	2.87	20.00	3.59
2	14.21	4.81	14.93	4.70	19.29	8.71	18.29	4.91

A repeated measures ANOVA, with adjustments for the Greenhouse-Geisser effect, showed a significant main effect of Score, $F(2.24, 58.34) = 6.39$, $p < .005$, no significant main effect of Group, $F(0.748, 19.45) = 1.06$, $p = .312$, and no significant interaction of Score and Group, $F(2.24, 58.34) = 1.91$, $p = .153$.

Table Three: Balsall Common Reading Ages
(on NFER Reading Tests administered each Summer)
Mean Reading age (RA) in months

Group	RA 2000	RA 2001	RA 2002	RA 2003	Change 2000 – 2001	Change 2001 – 2002	Change 2002 - 2003	Change 2001 - 2003
Control	101.71 (n=7)	109.45 (n=11)	128.45 (n=11)	137.00 (n=11)	-	18.09	8.64	28.00
Experimental	93.27 (n=11)	98.50 (n=12)	119.33 (n=12)	134.50 (n=12)	5.91	20.83	15.17	36.50
Total	96.56	103.74	123.70	135.70	-	19.52	12.04	32.43

A repeated measures ANOVA on the Reading Ages in 2001, 2002 and 2003 showed a significant main effect of Reading Age, F(2, 42) = 45.85, *p* < .001, no significant main effect of Group, F(1, 21) = 0.76, *p* = .393, and no significant interaction of Score and Group, F(2, 42) = 0.88, *p* = .423.

Table Four: Balsall Common SATs Results
(SAT's are graded in terms of levels (digits and letters). Possible grades for level 2 include 2a, b, c and 2d, which were coded as 2.75, 2.50, 2.25, 2.00, 1.90, with the same applying for equivalent levels 3, 4 and 5.)

Mean Maths Levels

Group	Maths 2000	Maths 2001	Maths 2002	Maths 2003	Change 2000 – 2001	Change 2001 – 2002	Change 2002 - 2003	Change 2001 - 2003
Control	-	3.34 (n=11)	3.87 (n=11)	4.25 (n=11)	-	0.53	0.38	0.91
Experimental	2.56 (n=11)	3.03 (n=12)	3.64 (n=12)	3.85 (n=12)	0.47	0.61	0.21	0.82
Total	2.56	3.18	3.75	4.04	-	0.57	0.29	0.86

A repeated measures ANOVA on Maths Scores for 2001, 2002 and 2003 showed a significant main effect of Score, F(2, 42) = 36.81, *p* < .001, no significant main effect of Group, F(1, 21) = 1.36, *p* = ..257, and no significant interaction of Score and Group, F(2, 42) = 0.34, *p* = .713.

Mean Comprehension Levels

Group	Comp 2000	Comp 2001	Comp 2002	Comp 2003	Change 2000 – 2001	Change 2001 – 2002	Change 2002 - 2003	Change 2001 - 2003
Control	-	3.06 (n=11)	3.62 (n=11)	4.43 (n=11)	-	0.56	0.81	1.37
Experimental	2.66 (n=11)	2.78 (n=12)	3.78 (n=12)	3.83 (n=12)	0.12	1.00	0.05	1.05
Total	2.66	2.91	3.70	4.10	-	0.79	0.40	1.19

A repeated measures ANOVA on Comprehension Scores for 2001, 2002 and 2003 showed a significant main effect of Score, $F(2,40) = 76.67$, $p < .001$, no significant main effect of Group, $F(1, 20) = 2.28$, $p = .147$, and a significant interaction of Score and Group, $F(2, 40) = 4.85$, $p < .05$.

Mean Writing Levels

Group	Writing 2000	Writing 2001	Writing 2002	Writing 2003	Change 2000 – 2001	Change 2001 – 2002	Change 2002 - 2003	Change 2001 - 2003
Control	-	2.44 (n=11)	3.04 (n=11)	3.38 (n=11)	-	0.60	0.34	0.94
Experimental	2.41 (n=11)	2.33 (n=12)	2.80 (n=12)	3.21 (n=12)	-0.08	0.47	0.41	0.88
Total	2.41	2.38	2.92	3.29	-	0.54	0.37	0.91

A repeated measures ANOVA on Writing Scores for 2001, 2002 and 2003 showed a significant main effect of Score, $F(2, 42) = 37.26$, $p < .001$, no significant main effect of Group, $F(1, 21) = 0.95$, $p = .341$, and no significant interaction of Score and Group, $F(2, 42) = 0.17$, $p = .841$.

<u>Table Five: Balsall Common Reading Ages and SATs by Risk Group</u>

Mean Reading Age in Months

Risk Group	Score 2000	*SD*	Score 2001	*SD*	Score 2002	*SD*	Score 2003	*SD*
1	100.1	19.74	108.3	21.23	128.4	31.79	136.4	22.43
2	92.1	13.25	98.82	13.45	118.5	21.26	134.9	21.24

A repeated measures ANOVA showed a significant main effect of Score, $F(2, 42) = 47.02$, $p < .001$, no significant main effect of Group, $F(1, 21) = 0.64$, $p = .432$, and no significant interaction of Score and Group, $F(2, 42) = 0.99$, $p = .380$.

Mean Maths Scores

Risk Group	Score 2000	*SD*	Score 2001	*SD*	Score 2002	*SD*	Score 2003	*SD*
1	2.75	0.59	3.44	0.60	4.06	0.62	4.49	3.59
2	2.40	0.33	2.90	0.59	3.41	0.57	3.59	0.71

A repeated measures ANOVA showed a significant main effect of Score, $F(2, 42) = 37.92$, $p < .001$, a significant main effect of Group, $F(1, 21) = 8.92$, $p < .01$, and no significant interaction of Score and Group, $F(2, 42) = 1.40$, $p = .258$.

Mean Comprehension Scores

Risk Group	Score 2000	*SD*	Score 2001	*SD*	Score 2002	*SD*	Score 2003	*SD*
1	2.75	0.50	3.12	0.66	3.85	0.66	4.17	0.69
2	2.58	0.52	2.69	0.46	3.55	0.67	4.03	0.67

A repeated measures ANOVA showed a significant main effect of Score, F(2, 40) = 64.79, p < .001, no significant main effect of Group, F(1, 20) = 0.71, p = .372, and no significant interaction of Score and Group, F(2, 40) = 0.70, p =.500.

Mean Writing Scores

Risk Group	Score 2000	*SD*	Score 2001	*SD*	Score 2002	*SD*	Score 2003	*SD*
1	2.45	0.21	2.43	0.49	3.15	0.49	3.37	0.65
2	2.38	0.14	2.32	0.28	2.66	0.57	3.20	0.35

A repeated measures ANOVA showed a significant main effect of Score, F(2, 42) = 40.08, p < .001, no significant main effect of Group, F(1, 21) = 2.22, p = .151, and no significant interaction of Score and Group, F(2, 42) = 1.96, p = .154.

Table Six: Balsall Common Treated Children Against the Whole school

Reading ages in months

Year	Weighted School Mean	Experimental Group Mean
2000	117.69	96.56 (N=18)
2001	130.94	103.74 (N=23)
2002	145.02	123.70 (N=23)
2003	152.84	134.86 (N=22)
Difference (2001-2003)	21.9	31.12

Maths SATs scores

Year	Weighted School Mean	Experimental Group Mean
2000	3.18	2.56 (N=11)
2001	3.72	3.18 (N=23)
2002	4.29	3.75 (N=23)
2003	4.46	4.00 (N=22)
Difference (2001-2003)	0.74	0.82

Comprehension SATs scores

Year	Weighted School Mean	Experimental Group Mean
2000	3.32	2.66 (N=11)
2001	3.70	2.91 (N=23)
2002	4.46	3.70 (N=23)
2003	4.70	4.06 (N=22)
Difference (2001-2003)	1.0	1.15

Writing SATs scores

Year	Weighted School Mean	Experimental Group Mean
2000	2.79	2.41 (N=11)
2001	2.97	2.38 (N=23)
2002	3.70	2.92 (N=23)
2003	3.76	3.26 (N=22)
Difference (2001-2003)	0.79	0.88

(Means for pupils across the relevant whole school year groups (minus pupils in experimental group) have been weighted across years, to correspond with the year group distribution of those in the experimental group. Not all pupils in the experimental group were administered tests in 2000, and in 2003 the eldest year group had passed out of the school. Calculation is of gain from 2001-2003, given that treatment began in Autumn 2001.)

Table Seven: Raw DST Effect Sizes for Time Four (2003) Versus Time Two (2001)

	Group 1 (Control then Experimental)	Group 2 (Experimental)
DST	2.09	2.45
Rapid Naming	1.54	1.39
Bead threading	1.25	1.39
One minute Reading	0.93	1.18
Postural Stability	4.67	3.63
Segmentation	1.48	1.85
Two minute Spelling	1.38	1.24
Backwards Span	1.03	1.06
Nonsense passage	1.58	1.38
One minute Writing	1.10	1.14
Verbal Fluency	0.44	0.71
Semantic Fluency	0.99	1.00

Table Eight: DST Decile Effect Sizes for Time Four (2003) Versus Time Two (2001)

	Group 1 (Control then Experimental)	Group 2 (Experimental)
Rapid Naming	0.49	0.66
Bead threading	0.56	0.65
One minute Reading	-0.11	0.20
Postural Stability	2.52	2.71
Segmentation	0.75	0.71
Two minute Spelling	0.17	0.13
Backwards Span	0.41	0.63
Nonsense passage	0.45	0.00
One minute Writing	0.14	0.23
Verbal Fluency	-0.21	0.11
Semantic Fluency	0.68	0.83