



National College for
Teaching & Leadership

Closing the gap: test and learn

Technical annex A: statistical analysis

Winter 2016

Contents

1	Statistical analysis of trial data	5
1.1	Preliminary assumption testing and the inferential tests used	5
1.2	Adjustment for design effect caused by cluster randomisation	6
2	Trial analysis results	8
2.1	1stClass@Number year 1 trial (all pupils) – ANCOVA and gain score analysis	9
2.2	1stClass@Number year 2 trial (all pupils) – ANCOVA and gain score analysis	11
2.3	1stClass@Number year 1 trial (FSM pupils) – ANCOVA and gain score analysis	13
2.4	1stClass@Number year 2 trial (FSM) – ANCOVA and gain score analysis	15
2.5	Growth Mindsets year 1 trial (all pupils) – ANCOVA and gain score analysis	17
2.6	Growth Mindsets year 2 trial (all new group reading test (NGRT)) – ANCOVA and gain score analysis	19
2.7	Growth Mindsets year 2 trial (all progress in maths (PiM)) – ANCOVA and gain score analysis	21
2.8	Growth Mindsets year 1 trial (FSM pupils) – ANCOVA and gain score analysis	23
2.9	Growth Mindsets year 2 trial (FSM NGRT) – ANCOVA and gain score analysis	25
2.10	Growth Mindsets year 2 trial (FSM PiM) – ANCOVA and gain score analysis	27
2.11	Inference Training year 1 trial (all pupils) – ANCOVA and gain score analysis	29
2.12	Inference Training year 1 trial (FSM pupils) – ANCOVA and gain score analysis	31
2.13	Numicon Intervention Programme year 1 trial (all pupils) – ANCOVA and gain score analysis	33
2.14	Numicon Intervention Programme year 2 trial (all pupils) – ANCOVA and gain score analysis	35
2.15	Numicon Intervention Programme year 1 trial (FSM pupils) – ANCOVA and gain score analysis	37
2.16	Numicon Intervention Programme year 2 trial (FSM pupils) – ANCOVA and gain score analysis	39

2.17	Research Lesson Study year 2 trial (all pupils) – ANCOVA and gain score analysis	41
2.18	Research Lesson Study year 2 trial (FSM pupils) – ANCOVA and gain score analysis	43
2.19	Response to Intervention (breakthroughs in literacy) (RTI) year 1 trial (all pupils) – ANCOVA and gain score analysis	45
2.20	RTI (breakthroughs in literacy) year 2 trial All – ANCOVA and Gain Score analysis	47
2.21	RTI (breakthroughs in literacy) year 1 trial (FSM pupils) – ANCOVA and Gain Score analysis	49
2.22	RTI (breakthroughs in literacy) year 1 trial (FSM) – ANCOVA and gain score analysis	51
2.23	RTI (breakthroughs in literacy) year 2 trial (writing, all) – ANCOVA and gain score analysis	53
2.24	RTI (breakthroughs in literacy) year 2 trial All pupils (two clusters removed) – ANCOVA and gain score analysis	55
2.25	RTI (breakthroughs in literacy) year 2 trial All pupils (multiple clusters removed) – ANCOVA and gain score analysis	57
2.26	RTI (breakthroughs in literacy) year 2 trial FSM pupils (two clusters removed) – ANCOVA and gain score analysis	59
2.27	RTI (breakthroughs in literacy) year 2 trial FSM pupils (multiple clusters removed) – ANCOVA and gain score analysis	61
2.28	Achievement for All trial year 1 (all pupils) – ANCOVA and gain score analysis	63
2.29	Achievement for All year 2 trial (all pupils) – ANCOVA and gain score analysis	65
2.30	Achievement for All year 1 trial (FSM) – ANCOVA and gain score analysis	67
2.31	Achievement for All year 2 trial (FSM pupils) – ANCOVA and gain score analysis	69
3	Pupil year groups involved in year 1 and year 2 trials	71
4	Attrition rates within the large-scale randomised controlled trials	72

1 Statistical analysis of trial data

1.1 Preliminary assumption testing and the inferential tests used

Prior to analysis, the principle was adopted that it is not acceptable to drop an outlier, just because it is an outlier – in order to use a parametric test. There was no evidence to suggest that any outliers were caused by data error, deliberate or accidental misreporting, sampling error or the non-maintenance of the research protocol.

Table 1 shows the hypotheses tested in each trial.

Table 1: Hypotheses

Experimental hypothesis 1a	Adjusting for pre-test scores, there will be an improvement in post-test scores for pupils exposed to the intervention for: (i) all pupils and (ii) FSM pupils
Experimental hypothesis 1b	Adjusting for both the design effect caused by cluster randomisation and pre-test scores, there will be an improvement in post-test scores for pupils exposed to the intervention for: (i) all pupils and (ii) FSM pupils
Experimental hypothesis 2a	There will be an improvement in the progress rates of pupils exposed to the intervention for: (i) all pupils and (ii) FSM pupils
Experimental hypothesis 2b	Adjusting for the design effect caused by cluster randomisation, there will be an improvement in the progress rates of pupils exposed to the intervention for: (i) all pupils and (ii) FSM pupils
Experimental hypothesis 3a	There will be a relationship between exposure to intervention and post-intervention test scores, taking into account pre-test scores, gender, age, FSM status, school Ofsted band, and proportion of FSM pupils in the school for: (i) all pupils and (ii) FSM pupils
Experimental hypothesis 3b	There will be a relationship between exposure to intervention and post-intervention test scores, taking into account pre-test scores, gender, age, FSM status, school Ofsted band, proportion of FSM pupils in the school and clustering of participants for: (i) all pupils and (ii) FSM pupils

With regard to hypotheses 1a (i–ii) and 1b (i–ii), with a view to using General Linear Model ANCOVA (analysis of co-variance) with pre-test scores as the covariate, preliminary assumption testing considered normality, the presence of outliers, homogeneity of

variance, homogeneity of regression slopes, linearity, homoscedasticity and equity of sample size between the control and intervention groups. None of the data met all of the assumptions necessary for use of this test. However, as the data met the assumption of similar distribution of co-variants, a non-parametric form of ANCOVA (Quade's F)¹ was able to be applied.

To test hypotheses 2a (i–ii) and 2b (i–ii), gain scores were first calculated from pre- and post-test scores; the gains for control and intervention were then assessed to see if they met the assumptions necessary for the use of one-way ANOVA (analysis of variance). Similarly, none of the data met all of the assumptions and therefore a non-parametric alternative (Kruskal-Wallis one-way ANOVA) was applied. The assumptions tested were: normality, the presence of outliers, homogeneity of variance, and equity of sample size.

Finally, in relation to hypotheses 3a (i–ii) and 3b (i–ii), the data was evaluated according to the requirements for conducting regression modelling with standard multiple regression. The majority of the results were satisfactory; however, two of the analyses had sample sizes that were too small for effective regression modelling to take place.

Using this variety of tests meant that the first two assessments were able to consider the effectiveness of the interventions with regard to the assessment of attainment, the second two relative progress, and the final pair of assessments were able to provide a validation of the other findings with regard to the effect of a range of individual differences at cluster and pupil level.

1.2 Adjustment for design effect caused by cluster randomisation

Adjustment for the design effect resulting from clustering in the testing of hypotheses 1b (i–ii) and 2b (i–ii) was carried out using the approach recommended and discussed by Campbell and colleagues.² This approach has been applied to primary healthcare where randomisation has had to take place at general practitioner surgery level. It involves using a formula to calculate the effective sample size taking into account the number of clusters and intracluster dependence. The p-value for the result is then adjusted accordingly.

Hypotheses 1a (i–ii) and 2a (i–ii) form an assessment of the effect of interventions without taking into account this design effect whilst hypotheses 1b (i–ii) and 2b (i–ii) represent adjustments to the results found in the preceding hypotheses. Interpretation of these two

¹ Quade, D. 'Rank analysis of covariance', *Journal of the American Statistical Association* 62 (1967): 1187–1200

² Campbell, M.K., Mollison, J., Steen, N., Grimshaw, J.M. and Eccles, M. 'Analysis of cluster randomized trials in primary care: a practical approach', *Family Practice* 17 (2000): 192–196. Available at: <http://fampra.oxfordjournals.org/content/17/2/192.long>.

sets of results allowed for an estimation of the effect of cluster randomisation across the whole programme of randomised controlled trials.

2 Trial analysis results

The following tables give details of the descriptive statistics, test statistics and probability values for each of the large-scale collaborative trials carried out during the programme.

2.1 1stClass@Number year 1 trial (all pupils) – ANCOVA and gain score analysis

Standard Age Scores (SAS)	Control		Intervention	
	Pre-test	Post-test	Pre-test	Post-test
n = 1139	565	565	574	574
Mean	89.75	90.93	91.93	93.69
Median	89.00	89.00	92.00	92.00
Standard deviation	13.37	14.18	13.56	13.73
Range	60.00	71.00	61.00	71.00
Standard error of the mean	0.56	0.60	0.57	0.57
Skewness	0.47	0.39	0.31	0.28
Kurtosis	-0.27	-0.43	-0.37	-0.47
95 per cent confidence interval (upper bound)	90.85	92.10	93.04	94.82
95 per cent confidence interval (lower bound)	88.65	89.76	90.82	92.57
Median gain in the control group				< 0.001
Median gain in the intervention group				< 0.001
Difference in gain				< 0.001
Design effect				2.08
Intracluster correlation coefficient				0.06
Effective sample size				558
Quade's non-parametric ANCOVA	F	p-value	n_p²	d*
Adjusted to compensate for the effect of cluster randomisation	1.35	0.245	0.002	0.099
Non-adjusted	2.81	0.094	0.002	0.099

One-way Kruskal-Wallis ANOVA on gain scores	X²	p-value	r**	d**
Adjusted to compensate for the effect of cluster randomisation	0.44	0.507	0.028	0.057
Non-adjusted	0.92	0.338	0.028	0.057

*Converted from F **Converted from chi-squared

Regression modelling	p-value
Taking into account pre-test scores, gender, age, free school meals (FSM) status, school Ofsted band, and proportion of FSM pupils in the school	0.010
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, proportion of FSM pupils in the school and clustering of participants	0.103

2.2 1stClass@Number year 2 trial (all pupils) – ANCOVA and gain score analysis

SAS	Control		Intervention	
	Pre-test	Post-test	Pre-test	Post-test
n = 513	250	250	263	263
Mean	83.59	85.17	85.29	87.21
Median	82.00	85.00	85.00	87.00
Standard deviation	10.42	11.62	11.36	12.47
Range	47.00	61.00	54.00	69.00
Standard error of the mean	0.66	0.74	0.70	0.77
Skewness	0.76	0.64	0.80	0.80
Kurtosis	0.32	0.35	0.70	0.76
95 per cent confidence interval (upper bound)	84.88	86.61	86.66	88.72
95 per cent confidence interval (lower bound)	82.30	83.73	83.92	85.70
<hr/>				
Median gain in the control group				3.00
Median gain in the intervention group				2.00
Difference in gain				-1.00
Design effect				1.42
Intracluster correlation coefficient				0.05
Effective sample size				369
<hr/>				
Quade's non-parametric ANCOVA	F	p-value	η_p^2	d*
Adjusted to compensate for the effect of cluster randomisation	0.53	0.466	0.001	0.077
Non-adjusted	0.75	0.385	0.001	0.077

One-way Kruskal-Wallis ANOVA on gain scores	χ^2	p-value	r**	d**
Adjusted to compensate for the effect of cluster randomisation	0.28	0.598	0.028	0.055
Non-adjusted	0.39	0.530	0.028	0.055

*Converted from F **Converted from chi-squared

Regression modelling	p-value
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, and proportion of FSM pupils in the school	0.147
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, proportion of FSM pupils in the school and clustering of participants	0.155

2.3 1stClass@Number year 1 trial (FSM pupils) – ANCOVA and gain score analysis

SAS	Control		Intervention	
	Pre-test	Post-test	Pre-test	Post-test
n = 257	143	143	114	114
Mean	83.06	84.84	87.98	89.47
Median	81.00	83.00	87.00	87.00
Standard deviation	12.32	12.88	12.15	13.18
Range	57.00	47.00	52.00	49.00
Standard error of the mean	1.03	1.08	1.14	1.23
Skewness	1.04	0.73	0.58	0.39
Kurtosis	0.99	-0.18	0.32	-0.72
95 per cent confidence interval (upper bound)	85.07	86.95	90.21	91.89
95 per cent confidence interval (lower bound)	81.04	82.73	85.75	87.05
Median gain in the control group				2.00
Median gain in the intervention group				<0.001
Difference in gain				-2.00
Design effect				1.09
Intracluster correlation coefficient				0.02
Effective sample size				240
Quade's non-parametric ANCOVA	F	p-value	η_p^2	d*
Adjusted to compensate for the effect of cluster randomisation	0.24	0.622	0.001	0.064
Non-adjusted	0.27	0.607	0.001	0.064

One-way Kruskal-Wallis ANOVA on gain scores	X²	p-value	r^{**}	d^{**}
Adjusted to compensate for the effect of cluster randomisation	0.03	0.871	0.011	0.021
Non-adjusted	0.03	0.866	0.011	0.021

*Converted from F **Converted from chi-squared

Regression modelling	p-value
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, and proportion of FSM pupils in the school	0.372
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, proportion of FSM pupils in the school and clustering of participants	0.369

2.4 1stClass@Number year 2 trial (FSM) – ANCOVA and gain score analysis

SAS	Control		Intervention	
	Pre-test	Post-test	Pre-test	Post-test
n = 132	62	62	70	70
Mean	84.27	86.52	83.83	85.99
Median	83.50	86.00	83.00	86.00
Standard deviation	11.14	12.87	11.09	12.00
Range	47.00	61.00	48.00	46.00
Standard error of the mean	1.41	1.63	1.33	1.43
Skewness	0.73	0.88	1.19	0.66
Kurtosis	0.13	1.13	1.84	-0.07
95 per cent confidence interval (upper bound)	87.05	89.72	86.43	88.80
95 per cent confidence interval (lower bound)	81.50	83.31	81.23	83.18
Median gain in the control group				2.50
Median gain in the intervention group				3.00
Difference in gain				0.50
Design effect				1.04
Intracluster correlation coefficient				0.02
Effective sample size				130
<hr/>				
Quade's non-parametric ANCOVA	F	p-value	η_p^2	d*
Adjusted to compensate for the effect of cluster randomisation	0.02	0.893	<0.001	0.024
Non-adjusted	0.02	0.891	<0.001	0.024

One-way Kruskal-Wallis ANOVA on gain scores	χ^2	p-value	r**	d**
Adjusted to compensate for the effect of cluster randomisation	0.16	0.691	0.035	0.071
Non-adjusted	0.16	0.685	0.035	0.071

*Converted from F **Converted from chi-squared

Regression modelling	p-value
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, and proportion of FSM pupils in the school	0.548
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, proportion of FSM pupils in the school and clustering of participants	0.661

2.5 Growth Mindsets year 1 trial (all pupils) – ANCOVA and gain score analysis

SAS	Control		Intervention		
	Pre-test	Post-test	Pre-test	Post-test	
n =	1772	850	850	922	922
Mean	100.74	102.53	99.58	100.06	
Median	101.00	103.00	100.00	101.00	
Standard deviation	16.29	17.55	14.82	16.68	
Range	68.00	72.00	72.00	72.00	
Standard error of the mean	0.56	0.60	0.49	0.55	
Skewness	-0.15	-0.18	-0.08	-0.00001	
Kurtosis	-0.72	-0.81	-0.62	-0.71	
95 per cent confidence interval (upper bound)	101.84	103.71	100.54	101.14	
95 per cent confidence interval (lower bound)	99.65	101.35	98.63	98.98	
Median gain in the control group					2.00
Median gain in the intervention group					1.00
Difference in gain					-1.00
Design effect					2.61
Intracluster correlation coefficient					0.05
Effective sample size					678
Quade's non-parametric ANCOVA	F	p-value	n_p^2	d^*	
Adjusted to compensate for the effect of cluster randomisation	4.93	0.027	0.007	-0.171	
Non-adjusted	12.88	< 0.001	0.007	-0.171	

One-way Kruskal-Wallis ANOVA on gain scores

	X²	p-value	r**	d**
Adjusted to compensate for the effect of cluster randomisation	2.62	0.105	- 0.062	- 0.125
Non-adjusted	6.86	0.009	- 0.062	- 0.125

*Converted from F **Converted from chi-squared

Regression modelling

	p-value
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, and proportion of FSM pupils in the school	0.000
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, proportion of FSM pupils in the school and clustering of participants	0.006

2.6 Growth Mindsets year 2 trial (all new group reading test (NGRT)) – ANCOVA and gain score analysis

SAS	Control		Intervention	
	Pre-test	Post-test	Pre-test	Post-test
n = 1195	618	618	577	577
Mean	94.00	95.25	93.90	95.87
Median	95.00	96.00	94.00	96.00
Standard deviation	14.64	15.48	15.29	16.20
Range	70.00	70.00	70.00	72.00
Standard error of the mean	0.59	0.62	0.64	0.67
Skewness	-0.08	0.01	0.16	0.17
Kurtosis	-0.70	-0.70	-0.56	-0.48
95 per cent confidence interval (upper bound)	95.16	96.47	95.15	97.19
95 per cent confidence interval (lower bound)	92.85	94.03	92.65	94.54
<hr/>				
Median gain in the control group				1.00
Median gain in the intervention group				2.00
Difference in gain				1.00
Design effect				3.80
Intracluster correlation coefficient				0.22
Effective sample size				498
<hr/>				
Quade's non-parametric ANCOVA	F	p-value	η_p^2	d*
Adjusted to compensate for the effect of cluster randomisation	0.38	0.539	0.001	0.069
Non-adjusted	1.43	0.232	0.001	0.069

One-way Kruskal-Wallis ANOVA on gain scores	χ^2	p-value	r**	d**
Adjusted to compensate for the effect of cluster randomisation	0.50	0.481	0.040	0.080
Non-adjusted	1.89	0.170	0.040	0.080

*Converted from F **Converted from chi-squared

Regression modelling	p-value
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, and proportion of FSM pupils in the school	0.473
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, proportion of FSM pupils in the school and clustering of participants	0.484

2.7 Growth Mindsets year 2 trial (all progress in maths (PiM)) – ANCOVA and gain score analysis

SAS	Control		Intervention	
	Pre-test	Post-test	Pre-test	Post-test
n = 998	518	518	480	480
Mean	90.24	91.74	89.74	91.49
Median	91.00	92.00	89.00	91.00
Standard deviation	13.45	12.02	13.74	12.65
Range	70.00	64.00	66.00	72.00
Standard error of the mean	0.59	0.53	0.63	0.58
Skewness	0.35	0.13	0.52	0.49
Kurtosis	-0.20	-0.14	0.04	0.63
95 per cent confidence interval (upper bound)	91.39	92.77	90.96	92.62
95 per cent confidence interval (lower bound)	89.08	90.70	88.51	90.36
<hr/>				
Median gain in the control group				1.00
Median gain in the intervention group				2.00
Difference in gain				1.00
Design effect				2.67
Intracluster correlation coefficient				0.13
Effective sample size				379
<hr/>				
Quade's non-parametric ANCOVA	F	p-value	η_p^2	d*
Adjusted to compensate for the effect of cluster randomisation	0.01	0.905	<0.001	0.012
Non-adjusted	0.04	0.845	<0.001	0.012

One-way Kruskal-Wallis ANOVA on gain scores	χ^2	p-value	r**	d**
Adjusted to compensate for the effect of cluster randomisation	0.002	0.962	0.002	0.005
Non-adjusted	0.006	0.938	0.002	0.005

*Converted from F **Converted from chi-squared

Regression modelling	p-value
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, and proportion of FSM pupils in the school	0.647
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, proportion of FSM pupils in the school and clustering of participants	0.695

2.8 Growth Mindsets year 1 trial (FSM pupils) – ANCOVA and gain score analysis

SAS	Control		Intervention	
	Pre-test	Post-test	Pre-test	Post-test
n = 441	203	203	238	238
Mean	92.77	94.14	92.62	94.18
Median	92.00	92.00	92.50	94.00
Standard deviation	15.30	14.90	14.89	15.90
Range	66.00	67.00	66.00	72.00
Standard error of the mean	1.07	1.05	0.97	1.03
Skewness	0.33	0.28	0.29	0.42
Kurtosis	-0.29	-0.44	-0.37	-0.32
95 per cent confidence interval (upper bound)	94.88	96.19	94.51	96.20
95 per cent confidence interval (lower bound)	90.67	92.09	90.73	92.16
Median gain in the control group				< 0.001
Median gain in the intervention group				1.50
Difference in gain				1.50
Design effect				1.25
Intracluster correlation coefficient				0.04
Effective sample size				360
Quade's non-parametric ANCOVA	F	p-value	η_p^2	d*
Adjusted to compensate for the effect of cluster randomisation	0.03	0.852	< 0.001	0.020
Non-adjusted	0.04	0.835	< 0.001	0.020

One-way Kruskal-Wallis ANOVA on gain scores

	χ^2	p-value	r**	d**
Adjusted to compensate for the effect of cluster randomisation	0.20	0.655	0.024	0.047
Non-adjusted	0.25	0.618	0.024	0.047

*Converted from F **Converted from chi-squared

Regression modelling

	p-value
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, and proportion of FSM pupils in the school	0.951
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, proportion of FSM pupils in the school and clustering of participants	0.956

2.9 Growth Mindsets year 2 trial (FSM NGRT) – ANCOVA and gain score analysis

SAS scores	Control		Intervention	
	Pre-test	Post-test	Pre-test	Post-test
n = 453	247	247	206	206
Mean	90.71	90.88	92.31	93.88
Median	91.00	90.00	93.00	93.00
Standard deviation	14.24	14.66	14.93	15.83
Range	58.00	70.00	70.00	72.00
Standard error of the mean	0.91	0.93	1.04	1.10
Skewness	0.14	0.26	0.34	0.34
Kurtosis	-0.82	-0.55	-0.28	-0.23
95 per cent confidence interval (upper bound)	92.48	92.71	94.35	96.04
95 per cent confidence interval (lower bound)	88.93	89.05	90.27	91.72
Median gain in the control group				-1.00
Median gain in the intervention group				<0.001
Difference in gain				1.00
Design effect				1.45
Intracluster correlation coefficient				0.10
Effective sample size				458
Quade's non-parametric ANCOVA	F	p-value	η_p^2	d*
Adjusted to compensate for the effect of cluster randomisation	2.35	0.126	0.008	0.173
Non-adjusted	3.40	0.066	0.008	0.173

One-way Kruskal-Wallis ANOVA on gain scores	χ^2	p-value	r**	d**
Adjusted to compensate for the effect of cluster randomisation	1.79	0.181	0.076	0.152
Non-adjusted	2.59	0.108	0.076	0.152

*Converted from F **Converted from chi-squared

Regression modelling	p-value
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, and proportion of FSM pupils in the school	0.328
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, proportion of FSM pupils in the school and clustering of participants	0.367

2.10 Growth Mindsets year 2 trial (FSM PiM) – ANCOVA and gain score analysis

SAS	Control		Intervention	
	Pre-test	Post-test	Pre-test	Post-test
n = 373	204	204	169	169
Mean	87.29	88.94	87.55	88.52
Median	87.00	90.00	87.00	88.00
Standard deviation	13.07	11.29	12.77	12.22
Range	61.00	50.00	52.00	60.00
Standard error of the mean	0.92	0.79	0.98	0.94
Skewness	0.50	-0.01	0.35	0.22
Kurtosis	-0.19	-0.40	-0.41	-0.43
95 per cent confidence interval (upper bound)	89.09	90.49	89.48	90.36
95 per cent confidence interval (lower bound)	85.50	87.39	85.63	86.68
Median gain in the control group				3.00
Median gain in the intervention group				1.00
Difference in gain				-2.00
Design effect				1.59
Intracluster correlation coefficient				0.12
Effective sample size				238
<hr/>				
Quade's non-parametric ANCOVA	F	p-value	n_p^2	d*
Adjusted to compensate for the effect of cluster randomisation	0.44	0.507	0.002	0.087
Non-adjusted	0.70	0.402	0.002	0.087

One-way Kruskal-Wallis ANOVA on gain scores	X²	p-value	r**	d**
Adjusted to compensate for the effect of cluster randomisation	0.90	0.343	0.062	0.124
Non-adjusted	1.43	0.231	0.062	0.124

*Converted from F **Converted from chi-squared

Regression modelling	p-value
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, and proportion of FSM pupils in the school	0.797
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, proportion of FSM pupils in the school and clustering of participants	0.825

2.11 Inference Training year 1 trial (all pupils) – ANCOVA and gain score analysis

SAS	Control		Intervention	
	Pre-test	Post-test	Pre-test	Post-test
n = 1477	875	875	602	602
Mean	95.16	97.53	96.23	98.55
Median	95.00	97.00	96.00	98.00
Standard deviation	14.17	14.44	15.86	16.47
Range	65.00	70.00	67.00	71.00
Standard error of the mean	0.48	0.49	0.65	0.67
Skewness	0.05	-0.01	0.11	0.03
Kurtosis	-0.71	-0.72	-0.79	-0.87
95 per cent confidence interval (upper bound)	96.10	98.49	97.50	99.87
95 per cent confidence interval (lower bound)	94.22	96.58	94.97	97.24
<hr/>				
Median gain in the control group				2.00
Median gain in the intervention group				2.00
Difference in gain				0.00
Design effect				2.88
Intracluster correlation coefficient				0.08
Effective sample size				521
<hr/>				
Quade's non-parametric ANCOVA	F	p-value	η_p^2	d^*
Adjusted to compensate for the effect of cluster randomisation	0.03	0.869	0.0001	-0.015
Non-adjusted	0.08	0.779	0.0001	-0.015

One-way Kruskal-Wallis ANOVA on gain scores

	χ^2	p-value	r**	d**
Adjusted to compensate for the effect of cluster randomisation	0.08	0.778	-0.012	-0.025
Non-adjusted	0.23	0.632	-0.012	-0.025

*Converted from F **Converted from chi-squared

Regression modelling

	p-value
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, and proportion of FSM pupils in the school	0.511
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, proportion of FSM pupils in the school and clustering of participants	0.573

2.12 Inference Training year 1 trial (FSM pupils) – ANCOVA and gain score analysis

SAS	Control		Intervention	
	Pre-test	Post-test	Pre-test	Post-test
n = 375	225	225	150	150
Mean	88.70	91.53	92.19	93.31
Median	88.00	92.00	90.50	92.00
Standard deviation	13.26	13.47	13.69	14.24
Range	55.00	60.00	58.00	57.00
Standard error of the mean	0.88	0.90	1.12	1.16
Skewness	0.22	0.17	0.30	0.21
Kurtosis	-0.90	-0.55	-0.63	-0.69
95 per cent confidence interval (upper bound)	90.43	93.29	94.38	95.59
95 per cent confidence interval (lower bound)	86.97	89.77	90.00	91.03
Median gain in the control group				4.00
Median gain in the intervention group				1.50
Difference in gain				0.00
Design effect				1.18
Intracluster correlation coefficient				0.03
Effective sample size				323
Quade's non-parametric ANCOVA	F	p-value	η_p^2	d^*
Adjusted to compensate for the effect of cluster randomisation	1.04	0.310	0.003	-0.114
Non-adjusted	1.22	0.269	0.003	0.114

One-way Kruskal-Wallis ANOVA on gain scores	χ^2	p-value	r**	d**
Adjusted to compensate for the effect of cluster randomisation	2.97	0.085	-0.097	-0.194
Non-adjusted	3.51	0.061	-0.097	-0.194

*Converted from F **Converted from chi-squared

Regression modelling	p-value
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, and proportion of FSM pupils in the school	0.474
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, proportion of FSM pupils in the school and clustering of participants	0.419

2.13 Numicon Intervention Programme year 1 trial (all pupils) – ANCOVA and gain score analysis

SAS scores	Control		Intervention	
	Pre-test	Post-test	Pre-test	Post-test
n = 875	371	371	504	504
Mean	92.10	93.34	88.13	90.72
Median	90.00	92.00	87.00	89.00
Standard deviation	13.13	14.18	13.75	14.39
Range	72.00	72.00	72.00	71.00
Standard error of the mean	0.68	0.74	0.61	0.64
Skewness	0.68	0.41	0.76	0.49
Kurtosis	0.64	0.15	0.41	-0.30
95 per cent confidence interval (upper bound)	93.43	94.78	89.33	91.98
95 per cent confidence interval (lower bound)	90.76	91.90	86.93	89.46
Median gain in the control group				2.00
Median gain in the intervention group				2.00
Difference in gain				< 0.001
Design effect				1.97
Intracluster correlation coefficient				0.05
Effective sample size				455
Quade's non-parametric ANCOVA	F	p-value	η_p^2	d*
Adjusted to compensate for the effect of cluster randomisation	0.61	0.435	0.001	0.074
Non-adjusted	1.20	0.273	0.001	0.074

One-way Kruskal-Wallis ANOVA on gain scores	χ^2	p-value	r**	d**
Adjusted to compensate for the effect of cluster randomisation	2.89	0.089	0.081	0.162
Non-adjusted	5.70	0.017	0.081	0.162

*Converted from F **Converted from chi-squared

Regression modelling	p-value
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, and proportion of FSM pupils in the school	0.111
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, proportion of FSM pupils in the school and clustering of participants	0.364

2.14 Numicon Intervention Programme year 2 trial (all pupils) – ANCOVA and gain score analysis

SAS	Control		Intervention	
	Pre-test	Post-test	Pre-test	Post-test
n = 529	278	278	251	251
Mean	88.69	90.77	86.91	96.22
Median	87.50	91.00	86.00	97.00
Standard deviation	14.24	13.75	13.94	14.22
Range	59.00	58.00	58.00	70.00
Standard error of the mean	0.85	0.82	0.88	0.90
Skewness	0.53	0.17	0.66	0.15
Kurtosis	-0.33	-0.85	-0.13	-0.29
95 per cent confidence interval (upper bound)	90.37	92.38	88.64	97.97
95 per cent confidence interval (lower bound)	87.02	89.15	85.19	94.46
<hr/>				
Median gain in the control group				3.50
Median gain in the intervention group				11.00
Difference in gain				7.50
Design effect				1.43
Intracluster correlation coefficient				0.04
Effective sample size				378
<hr/>				
Quade's non-parametric ANCOVA	F	p-value	η_p^2	d*
Adjusted to compensate for the effect of cluster randomisation	13.60	< 0.001	0.037	0.384
Non-adjusted	19.44	< 0.001	0.037	0.384

One-way Kruskal-Wallis ANOVA on gain scores	χ^2	p-value	r**	d**
Adjusted to compensate for the effect of cluster randomisation	13.32	< 0.001	0.190	0.386
Non-adjusted	19.04	< 0.001	0.190	0.386

*Converted from F **Converted from chi-squared

Regression modelling	p-value
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, and proportion of FSM pupils in the school	0.000
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, proportion of FSM pupils in the school and clustering of participants	0.000

2.15 Numicon Intervention Programme year 1 trial (FSM pupils) – ANCOVA and gain score analysis

SAS	Control		Intervention	
	Pre-test	Post-test	Pre-test	Post-test
n = 212	86	86	126	126
Mean	85.80	86.22	82.28	84.40
Median	85.00	85.50	81.00	84.00
Standard deviation	10.70	12.07	10.47	11.64
Range	45.00	46.00	48.00	59.00
Standard error of the mean	1.15	1.30	0.93	1.04
Skewness	0.38	0.29	0.84	0.91
Kurtosis	-0.06	-0.82	0.81	1.08
95 per cent confidence interval (upper bound)	88.06	88.77	84.11	86.44
95 per cent confidence interval (lower bound)	83.54	83.67	80.45	82.37
Median gain in the control group				0.50
Median gain in the intervention group				3.00
Difference in gain				2.50
Design effect				1.12
Intracluster correlation coefficient				0.03
Effective sample size				195
<hr/>				
Quade's non-parametric ANCOVA	F	p-value	η_p^2	d*
Adjusted to compensate for the effect of cluster randomisation	1.02	0.315	0.005	0.147
Non-adjusted	1.13	0.288	0.005	0.147

One-way Kruskal-Wallis ANOVA on gain scores	χ^2	p-value	r**	d**
Adjusted to compensate for the effect of cluster randomisation	2.14	0.144	0.106	0.213
Non-adjusted	2.38	0.123	0.106	0.213

*Converted from F **Converted from chi-squared

Regression modelling	p-value
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, and proportion of FSM pupils in the school	0.139
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, proportion of FSM pupils in the school and clustering of participants	0.219

2.16 Numicon Intervention Programme year 2 trial (FSM pupils) – ANCOVA and gain score analysis

SAS	Control		Intervention	
	Pre-test	Post-test	Pre-test	Post-test
n = 101	62	62	39	39
Mean	84.79	91.85	82.51	96.87
Median	82.50	91.00	82.00	97.00
Standard deviation	11.85	14.12	9.71	11.03
Range	43.00	50.00	31.00	50.00
Standard error of the mean	1.51	1.79	1.56	1.77
Skewness	0.66	0.23	0.23	0.16
Kurtosis	-0.40	-0.93	-1.10	0.19
95 per cent confidence interval (upper bound)	87.74	95.37	85.56	100.33
95 per cent confidence interval (lower bound)	81.84	88.34	79.46	93.41
Median gain in the control group				8.50
Median gain in the intervention group				15.00
Difference in gain				6.50
Design effect				1.03
Intracluster correlation coefficient				0.01
Effective sample size				102
<hr/>				
Quade's non-parametric ANCOVA	F	p-value	n_p^2	d*
Adjusted to compensate for the effect of cluster randomisation	6.20	0.014	0.065	0.506
Non-adjusted	6.39	0.013	0.065	0.506

One-way Kruskal-Wallis ANOVA on gain scores	X²	p-value	r^{**}	d^{**}
Adjusted to compensate for the effect of cluster randomisation	6.04	0.014	0.248	0.513
Non-adjusted	6.22	0.013	0.248	0.513

*Converted from F **Converted from chi-squared

Regression modelling	p-value
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, and proportion of FSM pupils in the school	0.216
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, proportion of FSM pupils in the school and clustering of participants	0.214

2.17 Research Lesson Study year 2 trial (all pupils) – ANCOVA and gain score analysis

SAS	Control		Intervention	
	Pre-test	Post-test	Pre-test	Post-test
n = 920	431	431	489	489
Mean	95.34	97.49	94.02	95.83
Median	95.00	97.00	94.00	96.00
Standard deviation	12.74	14.07	13.47	13.56
Range	60.00	68.00	64.00	71.00
Standard error of the mean	0.61	0.68	0.61	0.61
Skewness	-0.08	0.05	0.14	0.13
Kurtosis	-0.47	-0.38	-0.46	-0.47
95 per cent confidence interval (upper bound)	96.55	98.82	95.21	97.03
95 per cent confidence interval (lower bound)	94.14	96.16	92.82	94.63
Median gain in the control group				2.00
Median gain in the intervention group				2.00
Difference in gain				0.00
Design effect				1.96
Intracluster correlation coefficient				0.04
Effective sample size				483
Quade's non-parametric ANCOVA	F	p-value	η_p^2	d*
Adjusted to compensate for the effect of cluster randomisation	0.40	0.530	0.001	0.058
Non-adjusted	0.77	0.379	0.001	0.058

One-way Kruskal-Wallis ANOVA on gain scores	X²	p-value	r^{**}	d^{**}
Adjusted to compensate for the effect of cluster randomisation	0.37	0.544	0.028	0.056
Non-adjusted	0.72	0.396	0.028	0.056

*Converted from F **Converted from chi-squared

Regression modelling	p-value
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, and proportion of FSM pupils in the school	0.793
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, proportion of FSM pupils in the school and clustering of participants	0.767

2.18 Research Lesson Study year 2 trial (FSM pupils) – ANCOVA and gain score analysis

SAS	Control		Intervention	
	Pre-test	Post-test	Pre-test	Post-test
n = 163	75	75	88	88
Mean	91.68	93.96	91.11	93.56
Median	91.00	92.00	91.00	93.00
Standard deviation	11.94	13.87	13.80	14.35
Range	58.00	55.00	61.00	59.00
Standard error of the mean	1.38	1.60	1.47	1.53
Skewness	0.35	0.26	0.27	0.06
Kurtosis	-0.07	-0.74	-0.30	-0.61
95 per cent confidence interval (upper bound)	94.38	97.10	94.00	96.56
95 per cent confidence interval (lower bound)	88.98	90.82	88.23	90.56
Median gain in the control group				1.00
Median gain in the intervention group				2.00
Difference in gain				1.00
Design effect				1.06
Intracluster correlation coefficient				0.01
Effective sample size				159
Quade's non-parametric ANCOVA	F	p-value	n_p^2	d*
Adjusted to compensate for the effect of cluster randomisation	0.05	0.831	< 0.001	0.035
Non-adjusted	0.05	0.826	< 0.001	0.035

One-way Kruskal-Wallis ANOVA on gain scores	χ^2	p-value	r**	d**
Adjusted to compensate for the effect of cluster randomisation	0.04	0.841	0.016	0.032
Non-adjusted	0.04	0.836	0.016	0.032

*Converted from F **Converted from chi-squared

Regression modelling	p-value
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, and proportion of FSM pupils in the school	0.559
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, proportion of FSM pupils in the school and clustering of participants	0.581

2.19 Response to Intervention (breakthroughs in literacy) (RTI) year 1 trial (all pupils) – ANCOVA and gain score analysis

SAS	Control		Intervention	
	Pre-test	Post-test	Pre-test	Post-test
n = 1536	714	714	822	822
Mean	97.07	99.74	98.39	100.90
Median	98.00	100.00	99.00	102.00
Standard deviation	14.73	15.39	15.26	15.38
Range	69.00	72.00	70.00	72.00
Standard error of the mean	0.55	0.58	0.53	0.54
Skewness	0.07	-0.03	-0.05	-0.07
Kurtosis	-0.52	-0.65	-0.71	-0.74
95 per cent confidence interval (upper bound)	98.15	100.87	99.43	101.95
95 per cent confidence interval (lower bound)	95.99	98.61	97.34	99.85
<hr/>				
Median gain in the control group				2.00
Median gain in the intervention group				3.00
Difference in gain				1.00
Design effect				2.81
Intracluster correlation coefficient				0.07
Effective sample size				556
<hr/>				
Quade's non-parametric ANCOVA	F	p-value	η_p^2	d*
Adjusted to compensate for the effect of cluster randomisation	0.02	0.884	<0.001	0.012
Non-adjusted	0.06	0.807	<0.001	0.012

One-way Kruskal-Wallis ANOVA on gain scores	X²	p-value	r^{**}	d^{**}
Adjusted to compensate for the effect of cluster randomisation	0.18	0.668	0.018	0.037
Non-adjusted	0.52	0.472	0.018	0.037

*Converted from F **Converted from chi-squared

Regression modelling	p-value
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, and proportion of FSM pupils in the school	0.553
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, proportion of FSM pupils in the school and clustering of participants	0.664

2.20 RTI (breakthroughs in literacy) year 2 trial All – ANCOVA and Gain Score analysis

SAS	Control		Intervention	
	Pre-test	Post-test	Pre-test	Post-test
n = 317	157	157	160	160
Mean	86.80	90.36	86.45	88.70
Median	86.00	89.00	85.00	87.00
Standard deviation	12.05	12.66	11.77	13.60
Range	44.00	60.00	54.00	62.00
Standard error of the mean	0.96	1.01	0.93	1.08
Skewness	0.18	0.26	0.41	0.51
Kurtosis	-1.05	-0.62	-0.26	-0.35
95 per cent confidence interval (upper bound)	88.68	92.34	88.27	90.81
95 per cent confidence interval (lower bound)	84.91	88.38	84.63	86.59
<hr/>				
Median gain in the control group				3.00
Median gain in the intervention group				2.00
Difference in gain				-1.00
Design effect				1.55
Intracluster correlation coefficient				0.05
Effective sample size				213
<hr/>				
Quade's non-parametric ANCOVA	F	p-value	η_p^2	d*
Adjusted to compensate for the effect of cluster randomisation	1.45	0.230	0.007	-0.168
Non-adjusted	2.24	0.136	0.007	-0.168

One-way Kruskal-Wallis ANOVA on gain scores	χ^2	p-value	r**	d**
Adjusted to compensate for the effect of cluster randomisation	0.08	0.784	-0.019	-0.038
Non-adjusted	0.12	0.733	-0.019	-0.038

*Converted from F **Converted from chi-squared

Regression modelling	p-value
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, and proportion of FSM pupils in the school	0.326
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, proportion of FSM pupils in the school and clustering of participants	0.419

2.21 RTI (breakthroughs in literacy) year 1 trial (FSM pupils) – ANCOVA and Gain Score analysis

SAS	Control		Intervention	
	Pre-test	Post-test	Pre-test	Post-test
n = 460	248	248	212	212
Mean	93.18	96.10	91.38	93.15
Median	93.00	97.00	92.00	93.00
Standard deviation	13.86	14.48	14.12	15.23
Range	60.00	57.00	57.00	63.00
Standard error of the mean	0.88	0.92	0.97	1.05
Skewness	0.18	-0.03	0.10	0.19
Kurtosis	-0.65	-0.88	-0.85	-0.68
95 per cent confidence interval (upper bound)	94.90	97.90	93.28	95.20
95 per cent confidence interval (lower bound)	91.45	94.30	89.48	91.10
<hr/>				
Median gain in the control group				4.00
Median gain in the intervention group				1.00
Difference in gain				-3.00
Design effect				1.23
Intracluster correlation coefficient				0.03
Effective sample size				380
<hr/>				
Quade's non-parametric ANCOVA	F	p-value	η_p^2	d*
Adjusted to compensate for the effect of cluster randomisation	2.97	0.086	0.008	- 0.179
Non-adjusted	3.66	0.056	0.008	- 0.179

One-way Kruskal-Wallis ANOVA on gain scores

	χ^2	p-value	r**	d**
Adjusted to compensate for the effect of cluster randomisation	2.33	0.127	- 0.079	- 0.159
Non-adjusted	2.88	0.090	- 0.079	- 0.159

*Converted from F **Converted from chi-squared

Regression modelling

	p-value
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, and proportion of FSM pupils in the school	0.988
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, proportion of FSM pupils in the school and clustering of participants	0.987

2.22 RTI (breakthroughs in literacy) year 1 trial (FSM) – ANCOVA and gain score analysis

SAS	Control		Intervention	
	Pre-test	Post-test	Pre-test	Post-test
n = 116	57	57	59	59
Mean	84.26	89.00	83.83	85.02
Median	83.00	87.00	84.00	83.00
Standard deviation	11.39	12.47	10.03	12.28
Range	41.00	46.00	36.00	47.00
Standard error of the mean	1.51	1.65	1.31	1.60
Skewness	0.44	0.33	0.26	0.74
Kurtosis	-0.83	-0.66	-0.76	-0.01
95 per cent confidence interval (upper bound)	87.22	92.24	86.39	88.15
95 per cent confidence interval (lower bound)	81.31	85.76	81.27	81.88
<hr/>				
Median gain in the control group				4.00
Median gain in the intervention group				-1.00
Difference in gain				-5.00
Design effect				1.05
Intracluster correlation coefficient				0.01
Effective sample size				115
<hr/>				
Quade's non-parametric ANCOVA	F	p-value	η_p^2	d*
Adjusted to compensate for the effect of cluster randomisation	4.71	0.032	0.043	-0.414
Non-adjusted	4.92	0.028	0.043	-0.414

One-way Kruskal-Wallis ANOVA on gain scores

	χ^2	p-value	r**	d**
Adjusted to compensate for the effect of cluster randomisation	1.68	0.196	-0.123	-0.248
Non-adjusted	1.75	0.185	-0.123	-0.248

*Converted from F **Converted from chi-squared

Regression modelling

	p-value
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, and proportion of FSM pupils in the school	0.060
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, proportion of FSM pupils in the school and clustering of participants	0.176

2.23 RTI (breakthroughs in literacy) year 2 trial (writing, all) – ANCOVA and gain score analysis

SAS	Control		Intervention	
	Pre-test	Post-test	Pre-test	Post-test
n = 117	52	52	65	65
Median	54.00	62.50	60.00	56.00
Range	110.00	106.00	116.00	116.00
Median gain in the control group				8.50
Median gain in the intervention group				-4.00
Difference in gain				-12.50
Design effect				1.08
Intracluster correlation coefficient				0.01
Effective sample size				117
Quade's non-parametric ANCOVA				
	F	p-value	n_p^2	d*
Adjusted to compensate for the effect of cluster randomisation	0.003	0.959	< 0.001	0.010
Non-adjusted	0.003	0.958	< 0.001	0.010
One-way Kruskal-Wallis ANOVA on gain scores				
	X^2	p-value	r**	d**
Adjusted to compensate for the effect of cluster randomisation	0.002	0.964	0.004	0.009
Non-adjusted	0.002	0.963	0.004	0.009
*Converted from F	**Converted from chi-squared			

Regression modelling	p-value
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, and proportion of FSM pupils in the school	N/A
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, proportion of FSM pupils in the school and clustering of participants	N/A

2.24 RTI (breakthroughs in literacy) year 2 trial All pupils (two clusters removed) – ANCOVA and gain score analysis

SAS	Control		Intervention	
	Pre-test	Post-test	Pre-test	Post-test
n = 307	147	147	160	160
Mean	86.27	89.61	86.45	88.70
Median	86.00	89.00	85.00	87.00
Standard deviation	11.92	12.31	11.77	13.60
Range	44.00	46.00	54.00	62.00
Standard error of the mean	0.98	1.02	0.93	1.08
Skewness	0.23	0.22	0.41	0.51
Kurtosis	-0.98	-0.91	-0.26	-0.35
95 per cent confidence interval (upper bound)	88.19	91.60	88.27	90.81
95 per cent confidence interval (lower bound)	84.34	87.61	84.63	86.59
Median gain in the control group				3.00
Median gain in the intervention group				2.00
Difference in gain				-1.00
Design effect				1.58
Intracluster correlation coefficient				0.05
Effective sample size				202
Quade's non-parametric ANCOVA	F	p-value	n_p^2	d*
Adjusted to compensate for the effect of cluster randomisation	0.98	0.323	0.005	- 0.142
Non-adjusted	1.55	0.215	0.005	- 0.142

One-way Kruskal-Wallis ANOVA on gain scores	χ^2	p-value	r**	d**
Adjusted to compensate for the effect of cluster randomisation	0.08	0.777	- 0.020	- 0.041
Non-adjusted	0.13	0.722	- 0.020	- 0.041

*Converted from F **Converted from chi-squared

2.25 RTI (breakthroughs in literacy) year 2 trial All pupils (multiple clusters removed) – ANCOVA and gain score analysis

SAS	Control		Intervention	
	Pre-test	Post-test	Pre-test	Post-test
n = 237	114	114	123	123
Mean	86.19	88.06	86.02	88.89
Median	85.00	87.00	86.00	87.00
Standard deviation	11.67	13.33	12.04	12.16
Range	54.00	62.00	44.00	46.00
Standard error of the mean	1.09	1.25	1.09	1.10
Skewness	0.50	0.54	0.29	0.31
Kurtosis	-0.04	-0.16	-0.91	-0.86
95 per cent confidence interval (upper bound)	88.34	90.51	88.14	91.04
95 per cent confidence interval (lower bound)	84.05	85.61	83.89	86.75
Median gain in the control group				-2.00
Median gain in the intervention group				-1.00
Difference in gain				-1.00
Design effect				1.25
Intracluster correlation coefficient				0.03
Effective sample size				198
Quade's non-parametric ANCOVA	F	p-value	η_p^2	d*
Adjusted to compensate for the effect of cluster randomisation	0.64	0.423	0.003	-0.117
Non-adjusted	0.80	0.371	0.003	-0.117

One-way Kruskal-Wallis ANOVA on gain scores	χ^2	p-value	r**	d**
Adjusted to compensate for the effect of cluster randomisation	0.00	0.974	0.002	-0.005
Non-adjusted	0.00	0.971	0.002	-0.005

*Converted from F **Converted from chi-squared

2.26 RTI (breakthroughs in literacy) year 2 trial FSM pupils (two clusters removed) – ANCOVA and gain score analysis

SAS	Control		Intervention	
	Pre-test	Post-test	Pre-test	Post-test
n = 114	57	57	57	57
Mean	84.26	89.00	83.44	84.30
Median	83.00	87.00	83.00	82.00
Standard deviation	11.39	12.47	9.96	11.85
Range	41.00	46.00	36.00	47.00
Standard error of the mean	1.51	1.65	1.32	1.57
Skewness	0.44	0.33	0.33	0.84
Kurtosis	-0.83	-0.66	-0.65	0.42
95 per cent confidence interval (upper bound)	87.22	92.24	86.03	87.37
95 per cent confidence interval (lower bound)	81.31	85.76	80.85	81.22
Median gain in the control group				4.00
Median gain in the intervention group				-1.00
Difference in gain				-5.00
Design effect				1.06
Intracluster correlation coefficient				0.02
Effective sample size				113
Quade's non-parametric ANCOVA	F	p-value	η_p^2	d^*
Adjusted to compensate for the effect of cluster randomisation	5.61	0.020	0.053	- 0.458
Non-adjusted	5.92	0.017	0.053	- 0.458

One-way Kruskal-Wallis ANOVA on gain scores	χ^2	p-value	r**	d**
Adjusted to compensate for the effect of cluster randomisation	2.12	0.146	- 0.140	- 0.283
Non-adjusted	2.23	0.135	- 0.140	- 0.283

*Converted from F **Converted from chi-squared

2.27 RTI (breakthroughs in literacy) year 2 trial FSM pupils (multiple clusters removed) – ANCOVA and gain score analysis

SAS	Control		Intervention	
	Pre-test	Post-test	Pre-test	Post-test
n = 84	47	47	37	37
Mean	84.23	87.94	82.14	83.24
Median	82.00	86.00	82.00	81.00
Standard deviation	11.96	12.92	9.71	11.79
Range	41.00	46.00	35.00	46.00
Standard error of the mean	1.74	1.88	1.60	1.94
Skewness	0.44	0.41	0.29	0.89
Kurtosis	-0.92	-0.82	-0.85	0.59
95 per cent confidence interval (upper bound)	87.65	91.63	85.26	87.04
95 per cent confidence interval (lower bound)	80.82	84.24	79.01	79.44
Median gain in the control group				4.00
Median gain in the intervention group				-1.00
Difference in gain				-5.00
Design effect				1.02
Intracluster correlation coefficient				0.01
Effective sample size				86
Quade's non-parametric ANCOVA	F	p-value	η_p^2	d*
Adjusted to compensate for the effect of cluster randomisation	2.17	0.144	0.027	- 0.326
Non-adjusted	2.21	0.141	0.027	- 0.326
One-way Kruskal-Wallis ANOVA on gain scores	χ^2	p-value	r**	d**

Adjusted to compensate for the effect of cluster randomisation	0.14	0.704	- 0.042	- 0.084
Non-adjusted	0.15	0.701	- 0.042	- 0.084

*Converted from F **Converted from chi-squared

2.28 Achievement for All trial year 1 (all pupils) – ANCOVA and gain score analysis

SAS	Control		Intervention	
	Pre-test	Post-test	Pre-test	Post-test
n = 823	356	356	467	467
Mean	95.33	97.18	98.77	99.21
Median	95.00	96.00	100.00	101.00
Standard deviation	14.95	15.98	16.48	18.05
Range	67.00	69.00	72.00	72.00
Standard error of the mean	0.79	0.85	0.76	0.84
Skewness	0.10	0.15	-0.09	-0.04
Kurtosis	-0.59	-0.63	-0.73	-0.88
95 per cent confidence interval (upper bound)	96.88	98.84	100.27	100.85
95 per cent confidence interval (lower bound)	93.78	95.52	97.28	97.58
Median gain in the control group				1.00
Median gain in the intervention group				1.00
Difference in gain				0.00
Design effect				1.79
Intracluster correlation coefficient				0.03
Effective sample size				476
Quade's non-parametric ANCOVA	F	p-value	η_p^2	d*
Adjusted to compensate for the effect of cluster randomisation	1.76	0.186	0.004	-0.124
Non-adjusted	3.15	0.076	0.004	-0.124

One-way Kruskal-Wallis ANOVA on gain scores	χ^2	p-value	r**	d**
Adjusted to compensate for the effect of cluster randomisation	2.89	0.089	-0.079	-0.159
Non-adjusted	5.18	0.023	-0.079	-0.159

*Converted from F **Converted from chi-squared

Regression modelling	p-value
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, and proportion of FSM pupils in the school	0.261
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, proportion of FSM pupils in the school and clustering of participants	0.489

2.29 Achievement for All year 2 trial (all pupils) – ANCOVA and gain score analysis

SAS	Control		Intervention	
	Pre-test	Post-test	Pre-test	Post-test
n = 335	135	135	200	200
Mean	96.10	99.56	99.71	100.35
Median	95.00	97.00	102.50	103.00
Standard deviation	14.25	15.13	18.00	17.64
Range	58.00	70.00	69.00	69.00
Standard error of the mean	1.23	1.30	1.27	1.25
Skewness	0.27	0.28	-0.17	-0.20
Kurtosis	-0.57	-0.60	-1.03	-0.91
95 per cent confidence interval (upper bound)	98.51	102.12	102.20	102.79
95 per cent confidence interval (lower bound)	93.70	97.01	97.22	97.90
Median gain in the control group				2.00
Median gain in the intervention group				0.50
Difference in gain				-1.50
Design effect				1.27
Intraclass correlation coefficient				0.01
Effective sample size				280
Quade's non-parametric ANCOVA	F	p-value	η_p^2	d*
Adjusted to compensate for the effect of cluster randomisation	4.36	0.038	0.017	-0.258
Non-adjusted	5.54	0.019	0.017	-0.258

One-way Kruskal-Wallis ANOVA on gain scores

	χ^2	p-value	r**	d**
Adjusted to compensate for the effect of cluster randomisation	6.71	0.010	-0.160	-0.323
Non-adjusted	8.53	0.004	-0.160	-0.323

*Converted from F **Converted from chi-squared

Regression modelling**p-value**

Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, and proportion of FSM pupils in the school

N/A

Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, proportion of FSM pupils in the school and clustering of participants

N/A

2.30 Achievement for All year 1 trial (FSM) – ANCOVA and gain score analysis

SAS	Control		Intervention	
	Pre-test	Post-test	Pre-test	Post-test
n = 255	129	129	126	126
Mean	90.71	93.50	93.16	91.75
Median	92.00	93.00	93.00	89.00
Standard deviation	12.94	14.27	15.38	17.49
Range	55.00	55.00	53.00	62.00
Standard error of the mean	1.14	1.26	1.37	1.56
Skewness	0.16	0.10	-0.03	0.17
Kurtosis	-0.46	-0.57	-1.15	-1.30
95 per cent confidence interval (upper bound)	92.95	95.96	95.84	94.81
95 per cent confidence interval (lower bound)	88.48	91.03	90.47	88.70
Median gain in the control group				1.00
Median gain in the intervention group				-4.00
Difference in gain				-5.00
Design effect				1.16
Intracluster correlation coefficient				0.02
Effective sample size				228
Quade's non-parametric ANCOVA	F	p-value	η_p^2	d*
Adjusted to compensate for the effect of cluster randomisation	10.17	0.002	0.047	-0.431
Non-adjusted	11.82	< 0.001	0.047	-0.431

One-way Kruskal-Wallis ANOVA on gain scores	X²	p-value	r**	d**
Adjusted to compensate for the effect of cluster randomisation	12.98	< 0.001	-0.243	-0.501
Non-adjusted	15.08	< 0.001	-0.243	-0.501

*Converted from F **Converted from chi-squared

Regression modelling	p-value
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, and proportion of FSM pupils in the school	0.037
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, proportion of FSM pupils in the school and clustering of participants	0.060

2.31 Achievement for All year 2 trial (FSM pupils) – ANCOVA and gain score analysis

SAS	Control		Intervention	
	Pre-test	Post-test	Pre-test	Post-test
n = 100	41	41	59	59
Mean	94.56	95.80	94.64	95.68
Median	94.00	95.00	100.00	98.00
Standard deviation	12.89	14.32	19.13	17.99
Range	53.00	60.00	62.00	58.00
Standard error of the mean	2.01	2.24	2.49	2.34
Skewness	0.29	0.37	-0.08	-0.16
Kurtosis	-0.11	0.13	-1.41	-1.27
95 per cent confidence interval (upper bound)	98.51	100.19	99.52	100.27
95 per cent confidence interval (lower bound)	90.62	91.42	89.76	91.09
Median gain in the control group				1.00
Median gain in the intervention group				-2.00
Difference in gain				-3.00
Design effect				1.04
Intracluster correlation coefficient				0.01
Effective sample size				103
Quade's non-parametric ANCOVA	F	p-value	n_p^2	d*
Adjusted to compensate for the effect of cluster randomisation	0.003	0.953	0.00004	-0.012
Non-adjusted	0.004	0.953	0.00004	-0.012

One-way Kruskal-Wallis ANOVA on gain scores	X²	p-value	r^{**}	d^{**}
Adjusted to compensate for the effect of cluster randomisation	0.08	0.780	-0.028	-0.057
Non-adjusted	0.08	0.776	-0.028	-0.057

*Converted from F **Converted from chi-squared

Regression modelling	p-value
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, and proportion of FSM pupils in the school	N/A
Taking into account pre-test scores, gender, age, FSM status, school Ofsted band, proportion of FSM pupils in the school and clustering of participants	N/A

3 Pupil year groups involved in year 1 and year 2 trials

Intervention	Year 1	Year 2
1stClass@Number	3	3
Growth Mindsets	1, 3, 5, 7, 9	3, 5, 7, 9
Inference Training	3, 5, 7, 9	N/A
Numicon Intervention Programme	1, 3, 5	2
Response to Intervention (breakthroughs in literacy)	5, 7	5, 7
Research Lesson Study	1, 3, 5, 7, 9, 12	1, 3, 5, 7, 9, 12
Achievement for All	1, 3, 5, 7, 9, 12	

4 Attrition rates within the large-scale randomised controlled trials

Trials with randomisation at whole school level	Pre-test	Missing post-test	% of attrition
1stClass@Number	1910	613	32
Achievement for All (year 1)	1376	494	36
Achievement for All (year 2)	882	426	48
Growth Mindsets	3648	847	23
Inference Training	3006	1349	45
Numicon Intervention Programme	1368	324	24
RTI (Breakthroughs in Literacy)	3033	974	32
Totals	15223	5027	33

Year 2 replications with randomisation within each school	Pre-test	Missing post-test	% of attrition
1stClass@Number	898	247	28
Growth Mindsets (NGRT)	2183	405	19
Growth Mindsets (PiM)	1586	302	19
Numicon Intervention Programme	752	125	17
RTI (Breakthroughs in Literacy)	589	180	31
RTI (Breakthroughs in Literacy) (writing test)	476	314	66
Research Lesson Study	1108	188	17
Totals	7592	1761	23



National College for
Teaching & Leadership

© Crown copyright 2016

Reference: DFE-RR500c

ISBN: 978-1-78105-558-8

You may re-use this document/publication (not including logos) free of charge in any format or medium, under the terms of the Open Government Licence v3.0. To view this licence, visit www.nationalarchives.gov.uk/doc/open-government-licence/version/3 or email: psi@nationalarchives.gsi.gov.uk.

Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned.

The views expressed in this report are the authors' and do not necessarily reflect those of the Department for Education.

Any enquiries regarding this publication should be sent to us at: teaching.schools@education.gsi.gov.uk or www.gov.uk/nctl

This document is available for download at www.gov.uk/government/publications