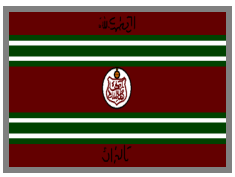
National Foundation for
Educational Research

Analysis of Hyderabad Data from 'Jolly Phonics' Initiative to Investigate its Impact on Pupil Progress in Reading and Spelling – India.



Statistics Research and Analysis Group (SRAG),
National Foundation for Educational Research (NFER),
United Kingdom.

Ian Schagen
Yarim Shamsan

Sponsor

Prof. James Tooley, Education Policy Professor,
University of Newcastle Upon Tyne,
United Kingdom.

27 June 2007

Analysis of Hyderabad Data from 'Jolly Phonics' Initiative to Investigate its Impact on Pupil Progress in Reading and Spelling – India.

Statistics Research and Analysis Group (SRAG), NFER, United Kingdom.

Ian Schagen
Yarim Shamsan

June 2007

Summary

Private schooling in the third world is on the increase and in particular in one of the low-income areas of India, where a data analysis on Hyderabad has been carried out[1].

The University of Newcastle has conducted an initiative research in Hyderabad on the impact on pupils' progress in Reading and Spelling tests. This initiative was called 'Jolly Phonics'.

The research was conducted in 20 private unaided English medium schools sited in the low-income areas of the city of Hyderabad. To carry out the programme, 265 children from 14 schools comprised the learning group and 241 children from 6 schools comprised the control group. The teaching and testing of the programme was carried out in 2004-2005.

This report outlines the analysis of examining the Jolly Phonics dataset in relation to two different groups, one is learning and the other is control.

The data comprises test scores as well as pupil background variables. Analysis of data will be carried out in the first instance. The main research questions will be explored and further supported by statistical evidence.
In what follows, a detailed analysis using appropriate statistical methodologies will be presented together with a final conclusion.

[1] Report via link: http://www.nfer.ac.uk/nfer/index.cfm?B98061C4-C299-53CD-A122-7C3A28E6B687.

Table of Contents

1. Introduction

The data had been collected in the low-income district of Hyderabad, shown on Map 1, to investigate the impact on the Jolly Phonics programme by the University of Newcastle. The Statistics Research and Analysis Group (SRAG) of the National Foundation for Educational Research (NFER) was commissioned to carry out a detailed analysis of this data in order to address a range of research issues relating to the apparent impact of such a programme on pupils' progress in reading and spelling tests.

Map 1 Hyderabad location on map[2]

Data on pupils' characteristics and tests results have been collected for 506 pupils in 20 schools taught by 5 different Jolly Phonics teachers.
Two school groups have been considered and studied: (1) learning and (2) control. The teaching of the programme started in August/September 2004 and ran for 6 months until February 2005. The lessons were for 1 hour everyday in the learning group schools carry out by trained teachers using Jolly Phonics materials.

The research issues to be explored included:

- What are the relationships between pupil characteristics and pupil outcomes in reading and spelling?
- What, if anything, is the impact of the Jolly Phonics initiative on pupils' progress in reading and spelling?
- Is the impact related to the number of days for which the pupil attended Jolly Phonics?
- Does the impact of Jolly Phonics vary for different kinds of pupil?

The main outcomes were scores on a number of tests in:

1. Burt reading test;
2. Schonell spelling test;
3. NFER[3] test A;

[2] Source: found @ https://www.cia.gov/cia/publications/factbook/print/in.html, accessed on 09/03/07.

4. NFER test B;
5. NFER test C;
6. Dictation test.

Each pupil took all six[4] tests at the start and end of the time period, and schools were allocated to receive or not receive Jolly Phonics instruction during the relevant period. Those who received instruction comprised the 'learning' group, while those who did not comprised the 'control' group. It should be noted that none of the control group appear to have done NFER B test at either time, so this test was excluded from the analysis.

In addition to test scores for each pupil at both time points, the following information was available for all or most pupils:

- Sex (boy/girl);
- Age (learning group only);
- IQ as measured by a coloured progressive matrices test;
- Number of days attended Jolly Phonics (learning group only).

From this data a set of models were set up in order to address the research questions outlined above.

2. Analysis

The analysis stages consisted of the following:

1. Initial data exploratory analysis,

2. Multilevel modelling then followed in order to address the principle research questions, but also to explore interactions between school types and other factors, and

3. Conclusion: Results have been reported in tables, graphs and brief summary of findings.

2.1 Exploratory Analysis

Descriptive analysis, basic frequencies and basic statistics were carried out for all the datasets' variables, including the number of missing values.

One of the exploratory analyses carried out is to examine the sex effect in school groups. Figure 1 shows an overview of the two different school groups by sex. It highlights that there are slightly more boys than girls in control group with 51% and 49% respectively. For the learning group, there are more girls than boys with 56% and 44% respectively.
Table 1 shows the number of pupils in each school group and the percentages of these.

[3] Note that these are nothing to do with NFER, and are presumably produced by the independent test publisher NFER-Nelson.
[4] With the exception of NFER B – see later.

Figure 1 School Groups by Sex

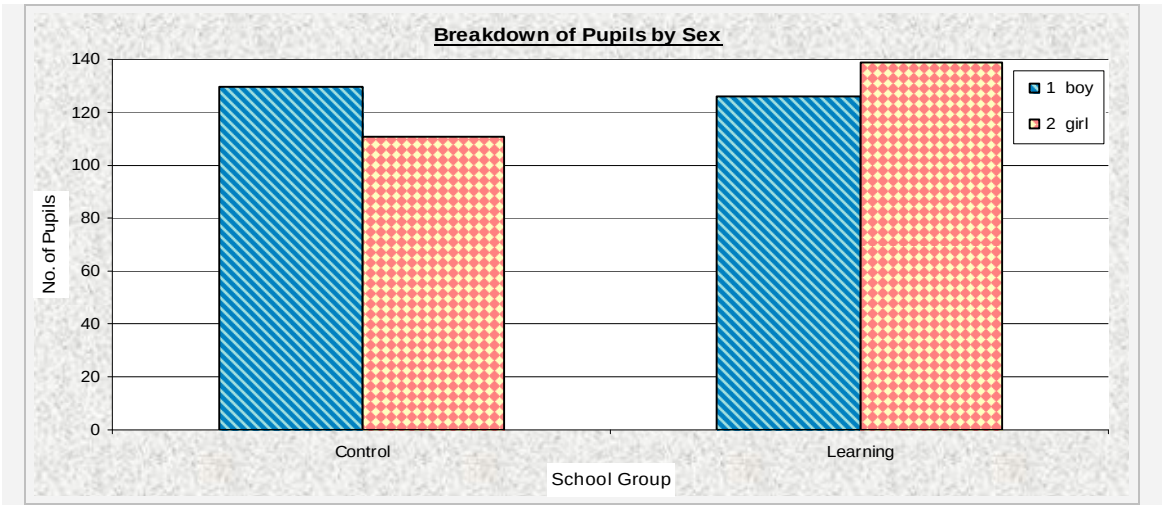


Table 1 School Groups Statistics

School Group	Freq.	Percent
Learning	241	48
Control	265	52

20 schools took part in the Jolly Phonics programme to represent either of two school groups. Figure 2 shows the breakdown of number of schools by learning and control school groups. It can be shown that mainly there were 2 schools with high number of pupils in the control group, namely school 10 and school 19. Furthermore, this can be illustrated in Figure 3 with a breakdown of the number of schools' by different school groups.

Figure 2 Schools breakdown by school groups

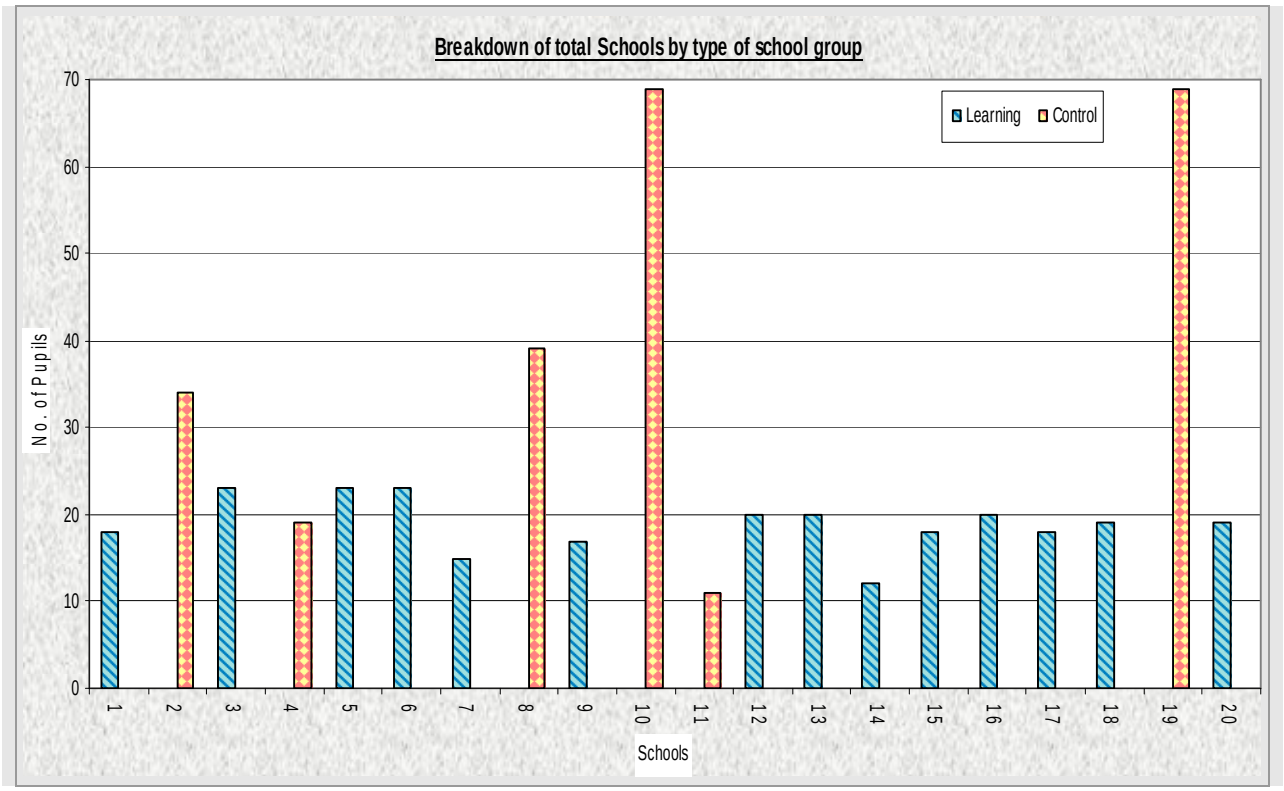
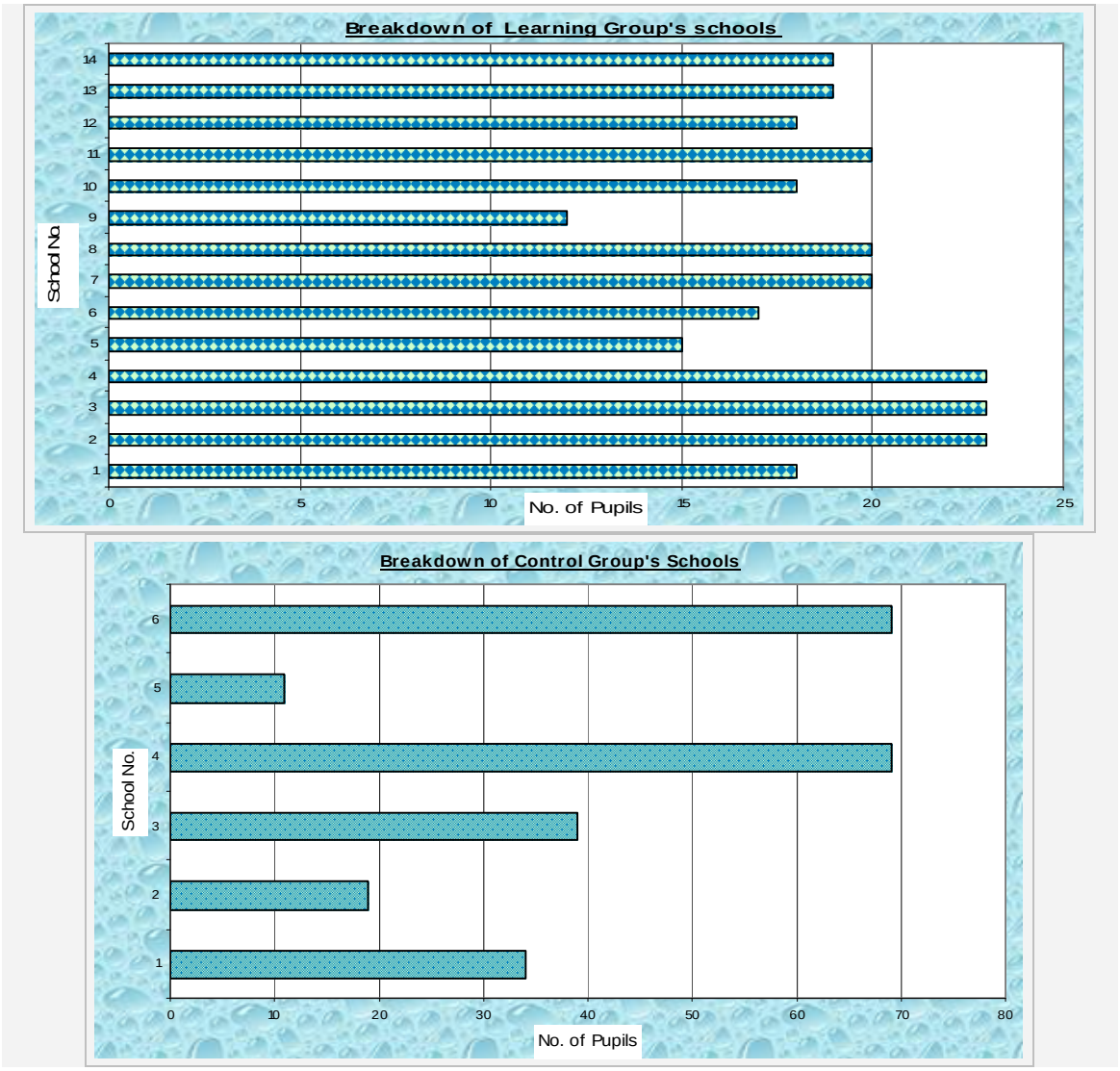
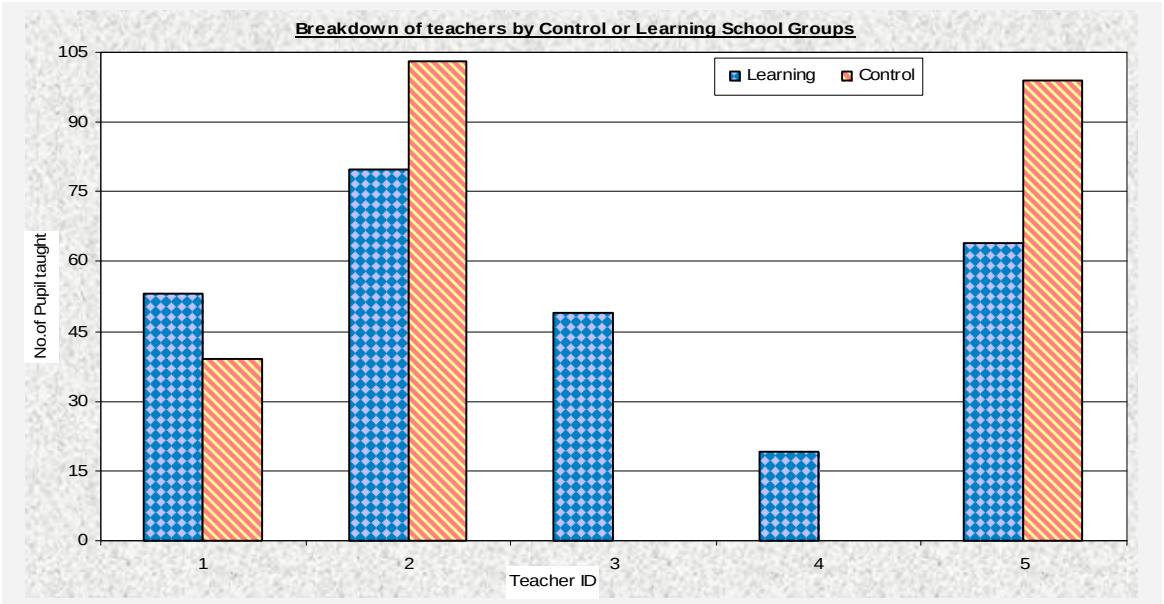


Figure 3 Breakdown of Learning and Control groups' schools



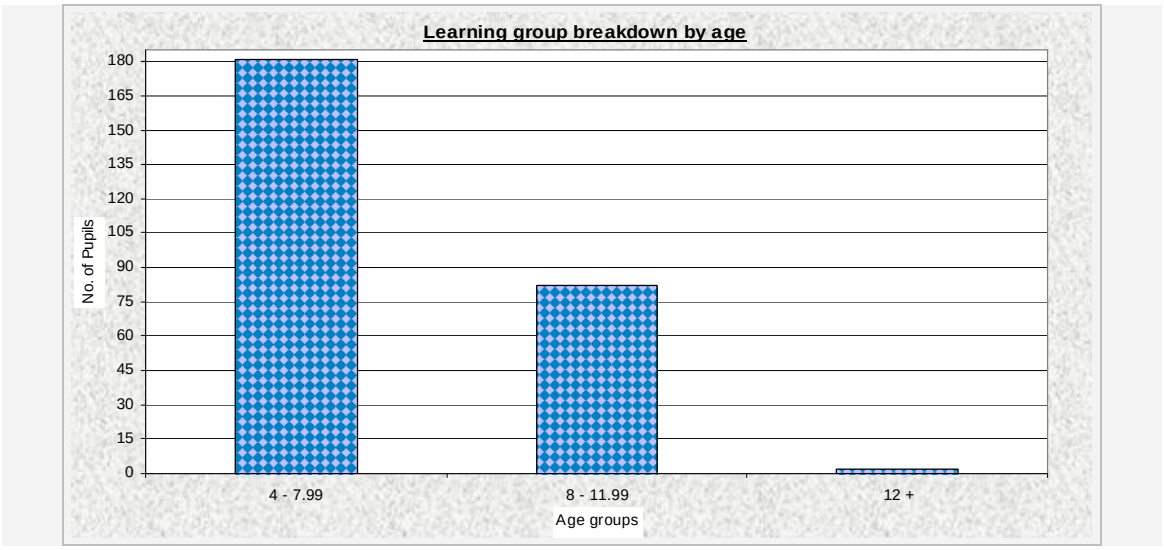
Five teachers were teaching the Jolly Phonics materials to the two school groups. Figure 4 shows the breakdown of teachers by the learning and the control group.

Figure 4 Breakdown of teachers by school groups



The breakdown of pupils' age groups to undertake the Jolly Phonics programme is illustrated in Figure 5. The programme was offered to under 14 years of age pupils in the learning group only. Generally, the learning group's under 8 years of age pupils were higher compared to the other age groups.

Figure 5 Breakdown of the Learning school group by Age groups



2.2 Multilevel Modelling

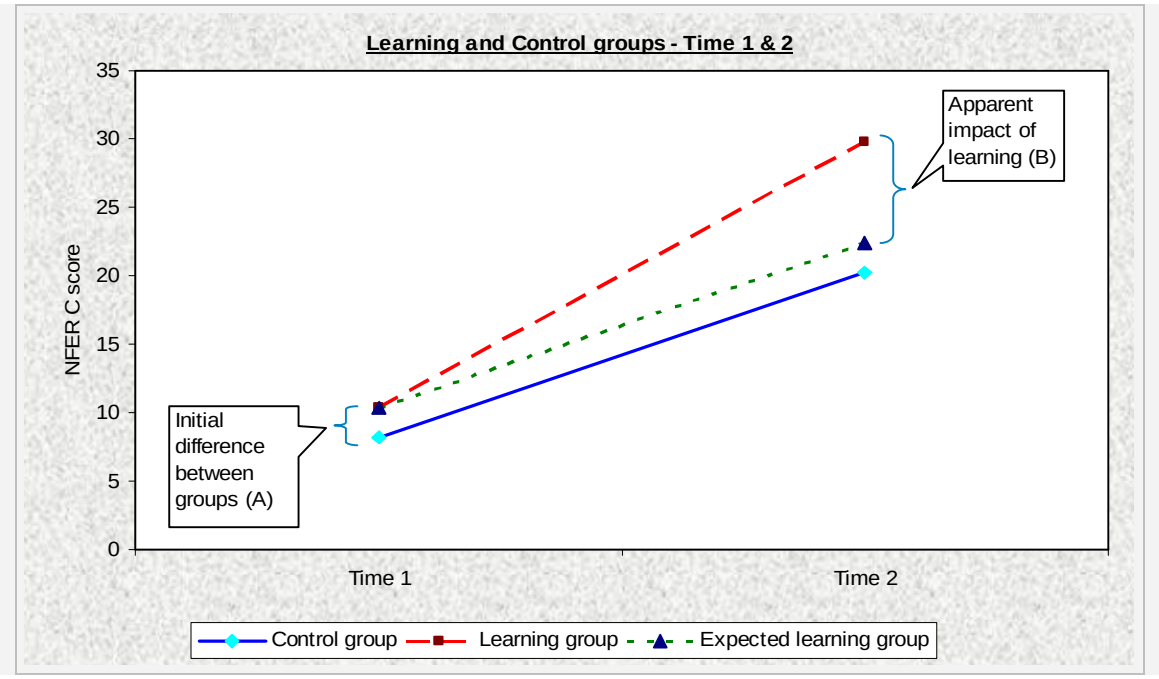
2.2.1 Background

Multilevel modelling (see Goldstein, 2003) is a development of regression analysis which takes account of data which is grouped into similar clusters at different levels. For example, individual pupils are grouped into classes, and those classes are grouped within schools. There may be more in common between pupils within the same class than with other classes, and there may be elements of similarity between different classes in the same school. Multilevel modelling allows us to take account of this hierarchical structure of the data and produce more accurate predictions, as well as estimates of the differences between pupils, between classes, and between schools.

2.2.2 Data Modelling

In this particular case there is an added complexity, in that each pupil has test scores at two time points – this is an example of a 'repeated measures' design. We assume there is an underlying 'trajectory of progress' from Time 1 to Time 2 (due to maturation, basic learning etc.) which is exemplified by the control group. For the learning group, we assume that there is an additional positive score change due to the effect of the learning occurring between the two time points. This model is illustrated simply in Figure 6 (using means scores for the NFER C test).

Figure 6 Illustration of Simple Repeated Measures Model



There are two elements of interest. One is labelled (A) and represents any overall difference between the two groups at Time 1 – the model assumes that without any learning effect this between-group difference would be the same at Time 2 (see the dotted line above labelled 'Expected learning group'). With the learning group, the dashed line represents the trajectory, and the difference labelled (B) is the apparent impact of the learning. It is this learning difference at Time 2 that we wish to estimate.

Each pupil has scores at both time points. These scores can be considered to comprise two parts: a deterministic part due to the learning and maturation effects as shown in Figure 6 as well as the pupil's own aptitude, and a random part due to 'noise' or measurement error. This latter 'noise' element can be modelled explicitly by including an extra level below the pupil to represent successive measurements on the same pupil controlling for other factors.

In total, the multilevel model was therefore set up with four levels:

1. Teacher of Jolly Phonics (5 teachers in total);
2. School (20 in total);
3. Pupil (506 in total);
4. Measurement occasion (2 of these per pupil).

The overall difference between learning and control groups (labelled (A) in Figure 6) was modelled by a group indicator which was the same (0 or 1) at both time points. The effect of the learning at Time 2 (labelled (B) in Figure 6) was modelled by an indicator which was only 1 for the learning group at Time 2 and 0 otherwise.

In order to address some of the research issues interaction terms were included in the models. These were created by multiplying together relevant variables in order to see if the coefficients of one variable were modified by the value of the other. For example, if we want to investigate if the impact of the learning at Time 2 was different for boys and girls, we would create an extra term in the model which was only 1 for girls in the learning group at Time 2 and 0 otherwise. This represents the additional effect of the learning on girls as opposed to boys. The variables for which such interactions were created were:

- Sex
- IQ
- Age

Annex 1 lists the variables that were included in the multilevel model set up as described above. The final models for each outcome are presented in Annexes 2 to 6, comprising random variances and fixed coefficients, with standard errors and 95% confidence intervals. These tables, however, are fairly hard to interpret as they stand, and give no information of the relative strength of the relationships between different factors and the outcomes. To overcome this, there is the usage of 'quasi effect sizes' (see Schagen, 2004). These enable us to present the results of complex models in a way which shows how much difference each factors makes to the expected pupil scores in each case. Annex 7 shows 'quasi effect sizes' for all the five outcomes and for background factors which are statistically significant at the 5% level. These are multiplied by 100 for clarity, but convey the average change in the outcome, expressed as a percentage of the outcome standard deviation, for an average change in the background variable.

Another interesting aspect of the results presented in Annexes 2 to 6 are the random variances, especially between teachers. These measure the differences in pupils' scores between the five different teachers, considered as a whole. None of these values are statistically significant, which implies that there is no evidence to suggest that the effect of Jolly Phonics varied according to the teacher involved.

The number of days for which pupils in the learning group attended Jolly Phonics programme was included in the dataset but is not presented in the main results summarised in Annexes 2 to 6. In modelling, this was confounded with the simple indicator of the learning effect, and thus perplexed the interpretation of the models. Annex 8 shows the values of the coefficients of this variable related to each of the outcomes, and the corresponding quasi effect sizes. There was only a significant effect for three of the outcomes.

2.2.3 Modelling conclusions

Based on these results, we may draw the following main conclusions from the modelling:

- There is a large and significant increase in all scores from Time 1 to Time 2, for both groups.
- Overall, girls tend to out-perform boys in the Burt and Schonell tests, and in NFER C.
- There appears to be no significant overall effect of age.
- Overall, there is no significant difference between learning and control groups at Time 1, except for the dictation score.
- All outcomes are strongly related to pupil's IQ, except for the dictation score.
- There is a significant apparent effect of the learning at Time 2 for all outcomes.
- For Schonell score, there is a stronger apparent effect of the learning group for girls than for boys.
- For all outcomes, the apparent effect of the learning group is stronger for younger pupils.
- There was no evidence that the learning effect was different for pupils with different IQs on any outcome.
- The number of days that pupils attended the Jolly Phonics programme appeared to have an impact on some outcomes.

3. Conclusions

Given the Jolly Phonics data and the analysis that has been carried out to answer the research issues, the following highlights the dataset's conclusions.

- What are the relationships between pupil characteristics and pupil outcomes in reading and spelling?

As might be expected, most outcomes were found to be significantly related to sex (girls out-performing boys) and to IQ. Interestingly, there was no evidence of a significant relationship with age. This may be because a lot of the data on age was missing[5], and set to the average.

- What, if anything, is the impact of the Jolly Phonics initiative on pupils' progress in reading and spelling?

The evidence from this analysis shows a clear indication of a positive impact of Jolly Phonics on test scores in reading and spelling, over and above the effect of maturation. The use of a learning and control groups design has enabled these two effects to be separated. Estimated effect sizes range from 0.26 to 0.62 (dividing values in Annex 7 by 100 to get values comparable to other studies), which are quite high for this type of intervention.

- Is the impact related to the number of days for which the pupil attended Jolly Phonics?

There is some evidence that the number of days attended had an impact on certain outcomes.

- Does the impact of Jolly Phonics vary for different kinds of pupil?

There is evidence to suggest that for one outcome (Schonell) the apparent impact was stronger for girls than boys. It also seemed to be stronger for younger pupils on all outcomes, but it is important to take into account the caveats about missing age data[6]. There was no evidence that the impact was different for pupils of different IQs.

Although the above findings from this data are clear and interesting, note should be taken of some caveats about over-interpretation of the results of the analysis. Statistical analysis can only control for factors which have been measured, and it also assumes that measurements in different settings are truly equivalent. It may be that there are other crucial factors which differ between the pupils undertaking Jolly Phonics programme and the others. Furthermore, the statistical analysis cannot make clear to us the reasons why the learning appears to be having the observed effect, and these reasons may be worth further investigation.

The analysis of the Jolly Phonics dataset has been fascinating with some expected results. The programme clearly shows its impact on various variables used in the research. As the analysis concluded, girls outperformed boys, where partners normally favour their boys to be more educated than girls. Thus, perhaps more attention should be directed to the partners in an 'awareness phase' of the Jolly Phonics programme.
Improvements can be expanded to include more schools, teachers and to gather more background characteristics on pupils and partners.
Hyderabad is one of the deprived areas of India, and the Jolly Phonics programme can be improved to offer a more worthy educational programme if it is to be followed.

[5] Note that the missing data was for the control group.
[6] As footnote 5.

4. Reference

Goldstein, H. (2003). *Multilevel Statistical Models.* 3rd ed. London: Arnold.

Schagen, I. (2004). 'Presenting the results of complex models - normalised coefficients, star wars plots and other ideas.' In: Schagen, I. and Elliot, K. (Eds) *But what does it mean? The use of effect sizes in educational research.* Slough: NFER.

Schagen, I. and Shamsan, Y. (2007) 'Analysis of International Data on the Impact of Private Schooling – Hyderabad, India', http://www.nfer.ac.uk/nfer/index.cfm?B98061C4-C299-53CD-A122-7C3A28E6B687.

Tooley, J and Dixon, P. (2003) Private Schools for the Poor: A case study from India, CfBt. Reading, Research and Development, http://www.ccs.in/ccsindia/pdf/psp_india.pdf.

Web References:

http://ncert.nic.in/sites/educationalsurvey/educationalsurvey.htm

http://en.wikipedia.org/wiki/Education_in_India#Structure

https://www.cia.gov/cia/publications/factbook/print/in.html

http://www.census.gov/cgi-bin/ipc/idbrank.pl

http://www.graphicmaps.com/citypops.htm

http://www.7thsurvey.ncert.nic.in/

http://www.worldsindhi.org/povertyreduction/21feb02.html

Annex 1 Variables available for analysis

Name	Min.	Max.	Description
teacher	1	5	Teacher id
school	1	20	School
pupil	1	506	Pupil ID
Time*	1	2	Time point
Sex	1	2	Sex (1 = boy, 2 = girl)
Age	4.8	13.9	Age (years)
Cons	1	1	Constant
learncon	0	1	Group (1 = treatment, 0 = control)
Iq	0	31	IQ (CPM)
learnt2*	0	1	Treatment at time 2
natt2*	0	126	Attendance (time 2)
Burtsco*	0	90	Burt score
Shonsco*	0	46	Schonell score
Nferasco*	0	10	NFER A score
Nferbsco*	0	36	NFER B score
Nfercsco*	0	34	NFER C score
Dictsco*	-9	20	Dictation score
Learnsex*	0	1	Treatment x sex interaction (time 2)
Learnage*	-2.6	6.5	Treatment by age interaction (time 2)
Learniq*	-9.7	16.3	Treatment by IQ interaction (time 2)

(* - value differs at Time 1 and Time 2)

Annex 2 Detailed Multilevel Model Results for *Burt Score*

				95% Confidence interval	
Parameter	**Estimate**	**Standard error**	**Sig.**	**Min.**	**Max.**
Base case					
Teacher variance	0.605	2.545		-4.383	5.594
School variance	9.804	4.699	*	0.594	19.014
Pupil variance	0.000	0.000		0.000	0.000
Time variance	144.800	6.528	*	132.005	157.595
Final model					
Teacher variance	2.971	3.526		-3.940	9.882
School variance	5.890	3.247		-0.474	12.254
Pupil variance	53.030	4.467	*	44.275	61.785
Time variance	29.680	1.879	*	25.997	33.363
Fixed coefficients					
Constant	-11.930	3.600	*	-18.986	-4.874
Time point	12.430	0.496	*	11.457	13.403
Sex (1 = boy, 2 = girl)	2.485	0.786	*	0.945	4.025
Age (years)	0.273	0.373		-0.459	1.005
Group (1 = treatment, 0 = control) **(A)**	1.457	1.524		-1.530	4.444
Treatment at time 2 **(B)**	4.729	0.845	*	3.072	6.386
IQ (CPM)	0.326	0.105	*	0.121	0.531
Treatment x sex interaction (time 2)	0.607	0.917		-1.191	2.404
Treatment by age interaction (time 2)	-0.732	0.340	*	-1.398	-0.065
Treatment by IQ interaction (time 2)	-0.041	0.116		-0.267	0.186

* - Estimate is significantly different from zero at the 5% level.

Annex 3 Detailed Multilevel Model Results for *Schonell Score*

				95% Confidence interval	
Parameter	**Estimate**	**Standard error**	**Sig.**	**Min.**	**Max.**
Base case					
Teacher variance	0.000	0.000		0.000	0.000
School variance	7.976	2.997	*	2.102	13.850
Pupil variance	0.000	0.000		0.000	0.000
Time variance	61.760	2.784	*	56.303	67.217
Final model					
Teacher variance	1.622	2.067		-2.429	5.673
School variance	4.545	2.108	*	0.413	8.677
Pupil variance	19.340	1.859	*	15.696	22.984
Time variance	16.670	1.055	*	14.602	18.738
Fixed coefficients					
Constant	-7.885	2.473	*	-12.732	-3.038
Time point	6.813	0.372	*	6.084	7.542
Sex (1 = boy, 2 = girl)	1.164	0.510	*	0.164	2.164
Age (years)	0.086	0.246		-0.396	0.567
Group (1 = treatment, 0 = control) **(A)**	1.496	1.232		-0.919	3.911
Treatment at time 2 **(B)**	4.292	0.630	*	3.058	5.526
IQ (CPM)	0.220	0.068	*	0.086	0.354
Treatment x sex interaction (time 2)	1.562	0.675	*	0.238	2.886
Treatment by age interaction (time 2)	-0.679	0.252	*	-1.173	-0.184
Treatment by IQ interaction (time 2)	-0.070	0.085		-0.237	0.097

* - Estimate is significantly different from zero at the 5% level.

Annex 4 Detailed Multilevel Model Results for *NFER A Score*

				95% Confidence interval	
Parameter	**Estimate**	**Standard error**	**Sig.**	**Min.**	**Max.**
Base case					
Teacher variance	0.060	0.091		-0.119	0.240
School variance	0.203	0.108		-0.010	0.415
Pupil variance	0.000	0.000		0.000	0.000
Time variance	4.305	0.194	*	3.925	4.685
Final model					
Teacher variance	0.048	0.073		-0.096	0.191
School variance	0.166	0.089		-0.008	0.340
Pupil variance	0.702	0.132	*	0.443	0.961
Time variance	2.125	0.134	*	1.862	2.388
Fixed coefficients					
Constant	-1.043	0.633		-2.283	0.197
Time point	1.988	0.133	*	1.728	2.248
Sex (1 = boy, 2 = girl)	0.045	0.135		-0.219	0.308
Age (years)	0.116	0.064		-0.010	0.242
Group (1 = treatment, 0 = control) **(A)**	0.077	0.261		-0.434	0.588
Treatment at time 2 **(B)**	0.562	0.221	*	0.129	0.994
IQ (CPM)	0.047	0.018	*	0.012	0.082
Treatment x sex interaction (time 2)	0.232	0.227		-0.214	0.677
Treatment by age interaction (time 2)	-0.195	0.087	*	-0.366	-0.025
Treatment by IQ interaction (time 2)	0.012	0.029		-0.044	0.069

* - Estimate is significantly different from zero at the 5% level.

Annex 5 Detailed Multilevel Model Results for *NFER C Score*

				95% Confidence interval	
Parameter	**Estimate**	**Standard error**	**Sig.**	**Min.**	**Max.**
Base case					
Teacher variance	0.000	0.000		0.000	0.000
School variance	12.780	5.030	*	2.921	22.639
Pupil variance	0.000	0.000		0.000	0.000
Time variance	127.800	5.761	*	116.508	139.092
Final model					
Teacher variance	3.046	3.289		-3.400	9.492
School variance	5.354	2.626	*	0.207	10.501
Pupil variance	24.790	2.840	*	19.224	30.356
Time variance	32.540	2.059	*	28.504	36.576
Fixed coefficients					
Constant	-12.280	3.051	*	-18.260	-6.300
Time point	12.040	0.520	*	11.021	13.059
Sex (1 = boy, 2 = girl)	2.147	0.631	*	0.910	3.384
Age (years)	0.211	0.304		-0.384	0.806
Group (1 = treatment, 0 = control) **(A)**	2.263	1.397		-0.475	5.001
Treatment at time 2 **(B)**	7.400	0.874	*	5.687	9.113
IQ (CPM)	0.250	0.084	*	0.084	0.415
Treatment x sex interaction (time 2)	0.004	0.925		-1.809	1.818
Treatment by age interaction (time 2)	-1.125	0.348	*	-1.807	-0.443
Treatment by IQ interaction (time 2)	0.060	0.117		-0.170	0.289

* - Estimate is significantly different from zero at the 5% level.

Annex 6 Detailed Multilevel Model Results for *Dictation Score*

				95% Confidence interval	
Parameter	**Estimate**	**Standard error**	**Sig.**	**Min.**	**Max.**
Base case					
Teacher variance	0.000	0.000		0.000	0.000
School variance	3.140	1.145	*	0.896	5.384
Pupil variance	0.000	0.000		0.000	0.000
Time variance	20.390	0.919	*	18.588	22.192
Final model					
Teacher variance	0.074	0.281		-0.475	0.624
School variance	1.015	0.507	*	0.020	2.010
Pupil variance	3.314	0.664	*	2.013	4.615
Time variance	10.930	0.691	*	9.575	12.285
Fixed coefficients					
Constant	-7.682	1.416	*	-10.457	-4.907
Time point	3.801	0.301	*	3.211	4.391
Sex (1 = boy, 2 = girl)	0.536	0.301		-0.054	1.125
Age (years)	0.172	0.144		-0.109	0.454
Group (1 = treatment, 0 = control) **(A)**	2.206	0.614	*	1.003	3.409
Treatment at time 2 **(B)**	1.598	0.500	*	0.619	2.577
IQ (CPM)	0.039	0.040		-0.039	0.117
Treatment x sex interaction (time 2)	0.309	0.513		-0.697	1.315
Treatment by age interaction (time 2)	-0.428	0.196	*	-0.812	-0.043
Treatment by IQ interaction (time 2)	0.102	0.065		-0.026	0.229

* - Estimate is significantly different from zero at the 5% level.

Annex 7 Summary of Significant Results Using Quasi Effect Sizes

Variable	Burt score	Schonell score	NFER A score	NFER C score	Dictation score
Time point	100	82	93	102	78
Sex (1 = boy, 2 = girl)	20	14		18	
Age (years)					
Group (1 = treatment, 0 = control) **(A)**					45
IQ (CPM)	15	16	13	12	
Treatment at time 2 **(B)**	38	51	26	62	33
Treatment x sex interaction (time 2)		19			
Treatment by age interaction (time 2)	-6	-8	-9	-10	-9
Treatment by IQ interaction (time 2)					

Annex 8 Coefficients of Days Attended Jolly Phonics programme (only for the learning group)

Outcome	Coefficient	SE	Sig	Quasi Effect size
Burt	0.0344	0.0305		4
Schonell	0.0821	0.0227	*	15
NFER A	0.0236	0.0075	*	16
NFER C	-0.0032	0.0311		0
Dictation	0.0739	0.0168	*	23

* - Estimate is significantly different from zero at the 5% level.

www.nfer.ac.uk

National Foundation for Educational Research
The Mere, Upton Park, Slough, Berkshire SL1 2DQ

Tel: 01753 574123
Fax: 01753 691623

For more information on:
- NFER, Statistics Research and Analysis Group: please see
http://www.nfer.ac.uk/about-nfer/departments/statistics-research-and-analysis-group.cfm
- E.G. West Centre in the School of Education at the University of Newcastle: please see
www.ncl.ac.uk/egwest