March 2005

The Third Implementation of the Sure Start Language Measure

Harris F, Law J and Roy P

City University, London

SureStart

Evidence & research

Research Report
SSU/2005/FR/014

The Third Implementation of the Sure Start Language Measure

Harris F, Law J and Roy P
City University, London

Glossary of terms

ANOVA
Analysis of Variance: A statistical technique to compare the mean scores of contrasting groups.

Chi-square
Statistical test which can compare the expected and observed values of data which are not assumed to be distributed normally.

Centiles
Centiles (also known as percentiles) indicate the value of a particular variable at a given point in the distribution. 25 % of the distribution lies at or below the 25^{th} centile, and so on.

Cohort
A set of cases (here, the data on a group of children).

Descriptive differences
Differences between groups of cases which can be described by graphical or numerical methods, but which do not reach statistical significance.

Distribution of scores
The graphical spread of scores, which plots the number of children for each of the possible scores on the word list.

Mean score
The arithmetical average of scores.

N
The total number of cases in a group

SSLM
Sure Start Language Measure

SSLP
Sure Start local programme

Sample
A group of cases (in this case, a group of children)

Sample source
The method of finding the sample from the population of the Sure Start local programme.

Standard deviation (sd)
This measures the degree of spread of scores away from the mean (average) score. A high standard deviation suggests a wide spread of scores around the mean.

Statistically significant
The difference between two or more groups is **significant** when there is a very low probability that the observed difference could be due to chance. Usually the probability level is set at 0.05.

Skew
This is a measure of the lean of a distribution graph either to the right or the left of its mean. A perfectly balanced distribution has an even spread of scores around its mean and zero skew.

Introduction

The Sure Start Language Measure (SSLM) is a parental report tool to measure change in the language skills of two year old children in Sure Start Communities. It was developed in 2001 for the Sure Start Unit by City University in conjunction with Sure Start local programmes. It has been designed to measure the Sure Start's PSA target on speech and language skills at both the national and local programme level. There have now been three national data collections: in 2001, in 2003 and 2004, known as the first, second and third implementations of the SSLM. Data will be collected twice more in 2005 and 2006.

In the first part of this report the full national SSLM data set of February – March 2004 is presented. In the second part, an analysis is made over all three implementations, using a substantial subset of the cases from each cohort. The report finishes with the conclusions about the comparisons and outlines future data collection plans. Appendices give greater technical background.

Key Findings

- There has been a statistically significant rise over the three years in the percentage of children with both high word count scores and parents without concerns about their child's language, from 70% to 76%.[1]

- Over the three years the *average* word count scores have been rising, although this does not reach a statistically significant level (26.2 words out of 50 in 2004, compared to 25.7 and 25.3 in 2003 and 2001 respectively)[2].

- The percentage of parents with language concerns about their child has dropped significantly [3]over the three years to 18% (from 20% and 22% in 2003 and 2001 respectively).

- The word count scores of boys have improved significantly over the three years (average word count score of 24 words out of 50 in 2004, compared to 23 and 22 in 2003 and 2001 respectively), while girls' average word count score is consistent at 28 words out of 50.[4]

- In 2004, English speaking children and bilingual children using English had an average word count score of 26 out of 50; while non-English speaking children had an average word count score of 25 out of 50.

- Over the three years, Sure Start registered children of all abilities have improved their scores, except for the children with a word count score of 10 or less.

- The proportion of children combining words into early sentences is stable (2004:84.1%, 2003: 84.4%, 2001: 84.6%).

[1] chi-square =14.02, df=3, p<0.05
[2] 2001 to 2004: F= 1.66, df 2,3632, p=0.191
[3] 2001 to 2004: Kruskal Wallis chi-square = 5.63, df 1, 2333, p= 0.018
[4] Boys only: F=3.858, df 2, 1859, p<0.05

The Sure Start Language Measure (SSLM)

The revised format of the SSLM for this third implementation consisted of a 50 word vocabulary checklist (reduced from 100 words) and four questions about parental concerns (reduced from 10 questions) including whether or not they have concerns about their child's speaking and listening. Both the original and the revised formats are presented in Appendices 1 and 2. The word list derives from that of the McArthur Communicative Development Inventory (MCDI) developed by Dale (Dale et al 2000). The shorter version is contained within the original format, making comparisons possible. Details of the SSLM revision process are being published separately (Harris et al, forthcoming).

As in previous implementations, the process of data collection followed published guidelines (Sure Start Unit 2003). Programmes were asked to sample two year old children from their programme geographical area. The SSLM questionnaire was used in a face-to-face interview with the child's main carer reporting on the child's spoken vocabulary.

Part 1 Third implementation data

1.1 The sample

1.1.1 Demographics

The number of children rose to 4694 (from 2866 in 2003 and 1615 in 2001.) 347 programmes participated in the third implementation, up from 227 (2003) and 127 (2001). Each programme submitted a mean of 13.5 cases, in line with previous years. The average age of each child was 23.9 months, with 51.5% being boys. There was an increase in the proportion of bilingual families to 26% of the sample (up from 20% in 2003, and 18% in 2001.) There was also a slight decrease in the proportion of special needs children, to 2.3 %, from 3% (in 2003). Third implementation sample descriptors are shown in Table 1.

Table 1: Sample descriptors for the third implementation (2004)

Population descriptors	**Third implementation**
Total number of children in the sample	4694
Average age of children [months]	23.9
Age range [range in months]	21-26
Gender [% male]	51.5
Family size [range]	1-11
Ordinal position	2.08
Bilingual children [%]	25.5
Children with special needs [%]	2.3
Parents with special needs [%]	2.4
Number of programmes	347
Cases per programme (average)	13.5 children
Main carer educated up to A /AS level [%]	45.6

A comparison of selected features of the samples across three years is shown in Table 2.

Table 2: Selected features of the three implementation (2004) samples

FULL DATA SETS	**2004**	2003	2001
Boys [%]	**51.5**	50.1	52.7
Language backgrounds [%] i. English only ii. Bilingual with English iii. Non-English speaking	**i. 74.5** **ii. 22.0** **iii. 3.6**	i. 79.9 ii. 16.6 iii. 3.5	i. 82 ii. 18 iii. na
Special needs children [%]	**2.3**	3.1	3.0

For the first time the education level of the main carer was sampled, as an additional descriptor of the cohort. The intention is to check future cohorts for comparability using measures of educational level and family size. Nearly 18% of the main carers are educated to degree level, while 20% are without qualifications. See Table 3. Here GCSE (General Certificate of Secondary Education) refers to the UK school examinations at the minimum school leaving age (any grade). (Equivalents from other education systems could be marked.) A/AS level refers to advanced level studies at secondary school after the minimum school leaving age.

Table 3: Educational level of main carer for the third implementation (2004)

Education level of main carer	Percentage of the sample	Number of cases
No qualifications	19.6	920
GCSE or equivalent	37.1	1743
A/AS Level	8.5	398
Degree level	17.6	826
Vocational skills	8.4	392
Not known	8.8	415
TOTAL	100	4694

The spread of expected education levels is shown in Table 4 using DfES statistics (DfES 2003). Cases without educational background information are removed for the purpose of comparison. The proportion of carers with no qualifications is high; this could well be related to the targeted geography of the Sure Start programmes. There are fewer carers educated to A/AS and degree level than the DfES statistics would expect. In part this is accounted for by those reporting vocational rather than academic skills.

Table 4: Educational levels of the third implementation (2004) dataset compared to UK statistics for 2003

Education level of main carer	Percentage of the sample without cases of unknown education background (n=4279)	DfES statistics 2003
No qualifications	21.5	14.1
GCSE or equivalent	40.7	41.9
A/AS Level	9.3	19.5
Degree level	19.3	24.6
Vocational skills	9.2	Not available
TOTAL	100	100

1.1.2 Sample sources

36% of the children were sampled in a representative[5] way (compared to 38% in 2003 and 39% in 2001.) A rising percentage of programmes (33%, up from 26 % in 2003 and 11% in 2001) used only children registered with their Sure Start programme for the SSLM data. (This could reflect an increase in reporting sample source). See Table 5.

Table 5: Sample sources for the three implementations

FULL DATA SETS	**2004**		***2003***		**2001**	
	Number of children	**Percent**	Number of children	Percent	Number of children	Percent
Representative	**1693**	**36.1**	1099	38.3	631	39.0
Wide sample	**599**	**12.8**	306	10.7	236	14.6
Sure Start only sample	**1535**	**32.7**	739	25.8	177	10.9
Mixed approaches	**287**	**6.1**	234	8.2	0	0
Unknown	**580**	**12.4**	488	17.0	573	35.4
Total	**4694**	**100**	2866	100.0	1617	100.0

Throughout the analysis checks were made for any effect of sample source. For this purpose the 'wide sample' and 'mixed approaches' categories were taken together, in order to look clearly at the contrast between representative sampling and sampling using only Sure Start registrations.

The profile of educational background varies only a little across the different sample sources (see Table 6.) There is most sample source variation in the cases of unknown educational background. When cases of unknown educational background are removed, there are similar educational profiles in each of the main sample groups of Table 6. This is confirmed by a Kruskal Wallis chi-square= 2.11, (df 3), p=0.55, not significant.

Compared to the remainder of the sample, the Sure Start only sample has slightly more vocational training and A/AS levels and GCSEs. However this could be accounted for by the lower percentage of cases with unknown educational backgrounds. The proportion of degree level education backgrounds is the same for the Sure Start sample type compared to the other sample types taken together.

[5] Representative: Children's names selected in an unbiased way from a comprehensive listing of two year olds
Wide sample: Children's names selected from various sources, including opportunistic contacts
Sure Start only: children's names taken from local programme registrations
Mixed approaches: combination of the above
Unknown: sample source not specified

Table 6: Sample types by educational background of main carer for the third implementation (2004)

SAMPLE SOURCE	**Representative**	**Sure Start**	Wide or mixed sample	Unknown
EDUCATIONAL LEVEL MAIN CARER	**N=1693**	**N=1535**	N=886	N=580
No qualifications	**19.6%**	**20.5%**	19.8%	17.1%
GCSE or equivalent	**36.4%**	**37.9%**	36.9%	37.6%
A/AS Level	**7.6%**	**9.6%**	8.8%	7.6%
Degree level	**16.0%**	**17.6%**	19.4%	19.5%
Vocational skills	**8.4%**	**9.3%**	6.9%	7.9%
Not known	**11.9%**	**5.2%**	8.2%	10.3%

1.2 Results for third implementation

1.2.1 Presentation of results

The findings presented here give key results for the word count score, word combining, and the level of parental concerns about language development. These results are then summarised into percentages that combine word count score with parental language concerns. Patterns of data are also presented by gender, by language backgrounds, by family educational background and by programme round number.

1.2.2 Key results

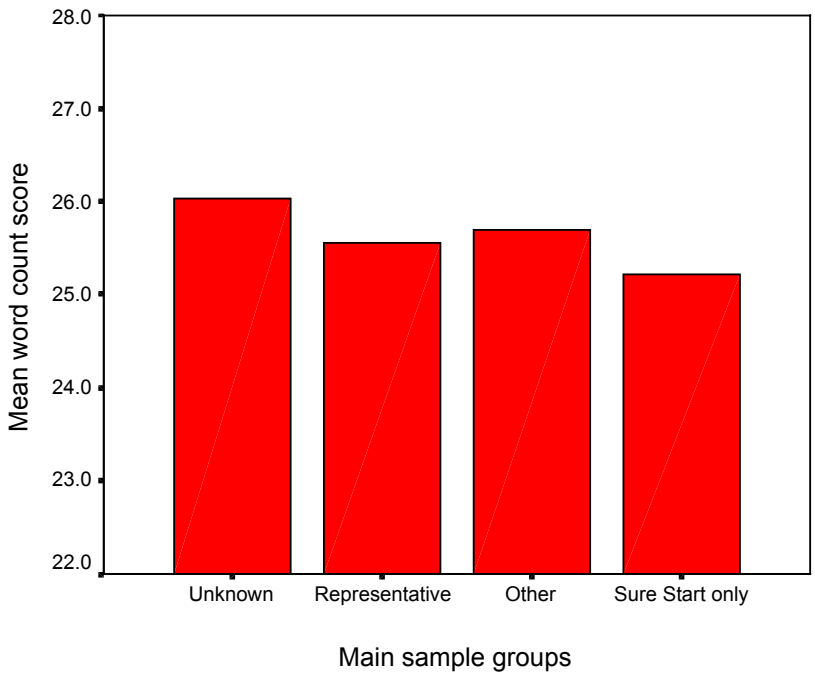
The findings here present the mean word count score out of 50, the rate of children combining words and the levels of parental concerns about their child's language (see Table 7). These results are then summarised into the summary percentages, which combine word counts and parental concerns.

Table 7: Main findings for the third implementation

	2004 FULL DATA SET N= 4694
Mean word count score (sd)	25.6 (12.6)
Children combining words Sometimes or often	84.1 %
Parents with language concerns	17.9 %

For each of these results, the mean word count score, the percentage of children combining words and the percentage of parents free of language concerns, there was no influence from the type of sampling procedure used by the programmes[6]. The mean word count scores by sample source are shown in Figure 1.

Figure 1: Mean word count score by main sample groups; third implementation (2004)



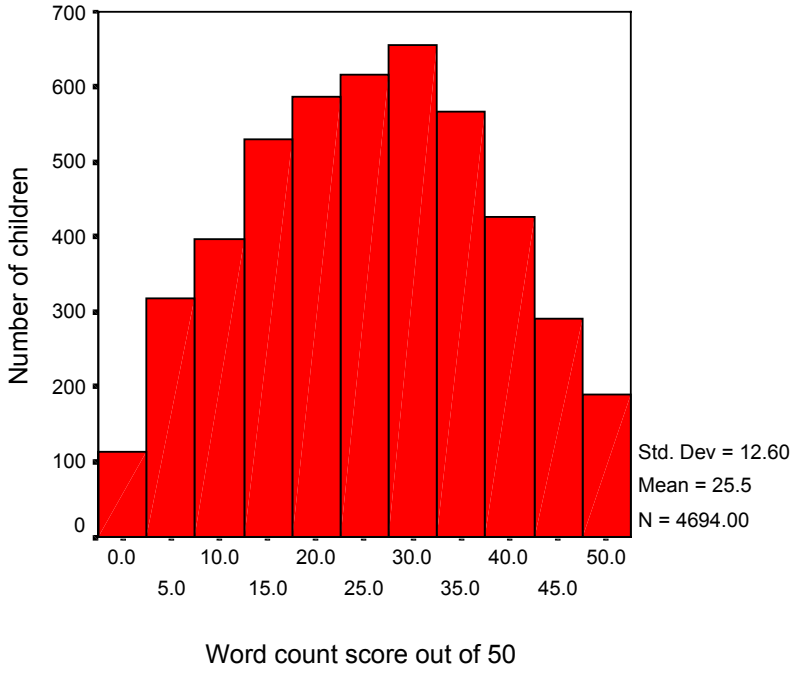
[6] Word count score by sample source: F = 0.686, df 3, 4690, p= 0.560.
Combining words by sample source: F = 0.25, df 3, 4683, p= 0.863.
Parental language concern by sample source: F = 0.472, df 3, 4689, p = 0.702.

1.2.3 Distribution of word count scores

Figure 2 shows the distribution of the word count scores.

The word count scores are not Normally distributed. This is confirmed by the Kolmogov-Smirnov statistic 0.041, df 4694, p<0.01. Instead of being symmetrical about its mean, the curve has a slight right-hand lean (or skew). In the two previous cohorts the curve was asymmetrical with a left hand lean.

Figure 2: Distribution of word count scores: third implementation (2004)



1.2.4 Summary figures

These key results can be summarised into percentages that combine word count score with parental language concerns. The percentages PQR and S are calculated by apportioning the cohort of children into four categories. The two parameters involved are word count score (above and below a threshold score) and presence or absence of parental concerns about language. The four categories are made up of the four possible combinations of these two parameters. The summary percentages PQR and S are used as indicators of change.

The four categories are shown in Table 8.

Table 8: Categories of children for calculating the summary percentages P Q R and S

	Low word count (12 or less)	High word count (13 or more)	TOTAL
Language concerns YES	A = Children with low word counts, AND parents concerned about their language development P%	B = Children with high word counts, AND parents concerned about their language development Q%	P+Q% = percentage of parents with language concerns
Language concerns NO	C = Children with low word counts, and parents *not* concerned about their language R%	D = Children with high word counts, and parents *not* concerned about their language development S%	R+S% = percentage of parents with no language concerns
TOTAL			P+Q+R+S=100 %

The number of children in each category is divided by the total number of children in the cohort (N) to give the summary percentages:

P= A/N x 100 %
Q=B/N x 100 %
R=C/N x 100 %
S=D/N x 100 %

Detailed analysis of nearly 2000 cases from the first and second implementations allows the level of 20 out of 100 on the original SSLM to be matched to a level of 12 out of 50 on the revised (shortened) SSLM. For more detail see Appendix 4. The summary figures for the third implementation are here calculated using a threshold score of 12 (see Tables 9 and 10). These percentages can be compared to the PQR and S figures already published for 2001 and 2003 for the original SSLM format.

Table 9: Word count scores and parental language concerns for the 2004 cohort

	Low word count (12 or less)	High word count (13 or more)	TOTAL
Language concerns YES	448	393	841
Language concerns NO	379	3472	3852
TOTAL	827	3866	4693[7]

[7] One case out of 4694 had missing language concern data

Table 10: Summary percentages for the third implementation (2004)

	Low word count (12 or less)	High word count (13 or more)	TOTAL
Language concerns YES	P= 9.5 %	Q= 8.4 %	17.9 %
Language concerns NO	R= 8.1 %	S= 74.0 %	82.1 %
TOTAL	17.6 %	82.4 %	100.0 %

1.2.5 Gender

Boys had a mean word count score of 23 out of 50, significantly lower than girls who had a mean word count score of 28. Boys were significantly less likely to be combining words into sentences (21% not combining words compared to 11% of girls) and their parents significantly more often expressed concerns about their son's language development (22% compared to 14%). For each variable, girls were significantly stronger than boys[8]. This is consistent with the two previous years' data.

Table 11: Comparison of boys and girls for the third implementation (2004)

	BOYS N= 2413	**GIRLS** N= 2272
Mean word count score (sd)	23.1 (12.7)	28.1 (12.0)
Children combining words Sometimes or often	79.4 %	89.1 %
Parents with language concerns	22.1 %	13.5 %

The graphs of the word count distributions are shown for boys and girls separately in Figures 3 and 4. Neither distribution is Normal, as evidenced by the Kolmogorov-Smirnov statistics. (Boys K-S=0.052, df 2413, p<0.001, girls K-S=0.048, df 2272, p<0.001.) The boys' distribution of scores lies with a left-hand cluster (positive standardised skew of 3.46) while the girls' distribution of scores lies with a right-hand cluster (negative standardised skew of –3.25.) Throughout the range of the centiles of the distribution, the girls' scores are higher than the boys.

[8] Word count score by gender: F = 186.5, df 1,4683, p< 0.01
Word combining by gender: F = 136.6, df 1,4676, p< 0.01.
Language concerns by gender: F = 59.1, df 1,4682, p<0.01.

Figure 3: Word count scores of boys; third implementation (2004)

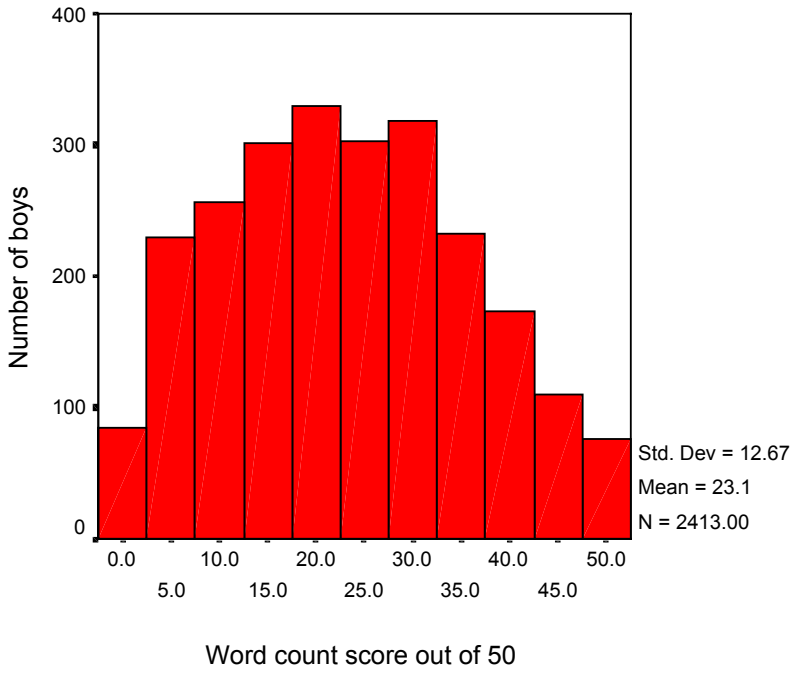
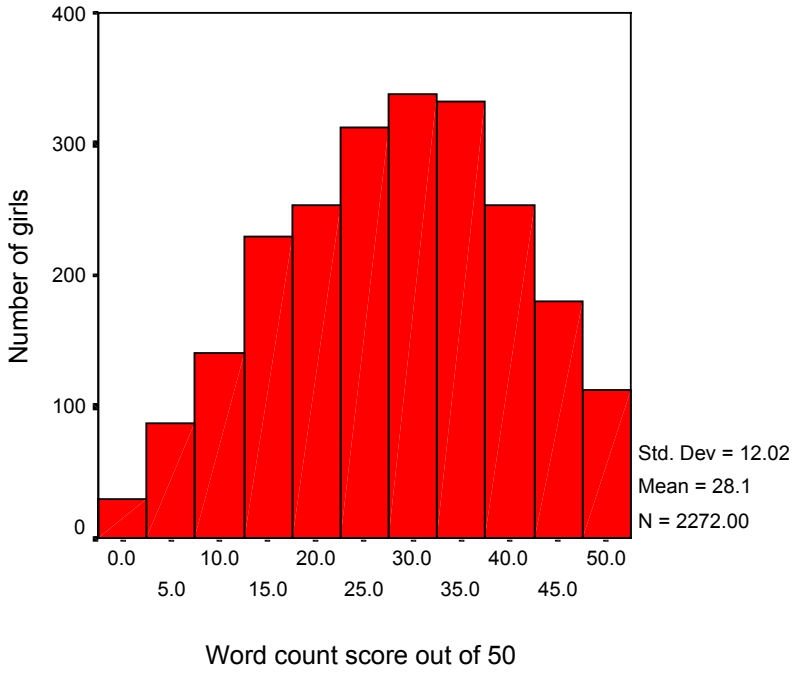


Figure 4: Word count scores of girls; third implementation (2004)



1.2.6 Language background

Table 12 shows the key results by the language background of the child.

Table 12: Third implementation (2004) results by language background

LANGUAGE BACKGROUND	ENGLISH ONLY	BILINGUAL WITH ENGLISH	NON-ENGLISH SPEAKING
Number of children (4691 cases)	3494	1032	165
Mean word count score (sd)	25.8 (12.7)	24.8 (12.2)	24.5 (13.3)
Children combining words Sometimes or often	85.0%	81.8%	79.3%
Parents with language concerns	18%	17%	18%

The 50 word format gives comparable results for word count score for the different language backgrounds, for the 4691 cases of known language background. The differences between these language backgrounds are not statistically significant: F= 2.941, df 2, 4688, p=0.053. This is a positive result, suggesting that the revised word list does not disadvantage children from varied language backgrounds. Earlier implementations showed a significantly higher mean word count score for children from English only backgrounds. This could have been due either to the word list itself disadvantaging the bilinguals, or to language learning differences in the bilingual populations. The comparability of results for the third implementation using the revised SSLM format suggests that the original word list may have been contributing to lower scores for the bilingual groups in earlier implementations.

The rate of combining words does however vary significantly according to language background, with more word combinations reported by the English-only parents. (Kruskal Wallis chi-square=39.43, df 2, $p<0.01$.) Further breakdown of the word combining data is given in Table 13.

Table 13: Word combining by language background; third implementation (2004)

LANGUAGE BACKGROUND	ENGLISH ONLY	BILINGUAL WITH ENGLISH	NON-ENGLISH SPEAKING
Not combining words (number of children)	15.0 % (523)	18.2 % (188)	20.6 % (34)
Sometimes combining words (number of children)	23.9 % (834)	30.4 % (314)	34.5 % (57)
Often combining words (number of children)	61.1 % (2130)	51.4 % (530)	44.8 % (74)
TOTALS			

The level of parental concerns about language does not vary significantly with language background (Kruskal Wallis chi-square = 0.157,df 2, p=0.925.) See Table 14.

Table 14: Parental language concerns by language background; third implementation (2004)

LANGUAGE BACKGROUND	ENGLISH ONLY	BILINGUAL WITH ENGLISH	NON-ENGLISH SPEAKING
No language concerns %	81.9	82.7	81.8
Mild concern about speaking OR listening %	11.4	9.2	8.5
Some concern about speaking OR listening %	5.4	6.2	7.9
Concerned about speaking AND listening %	0.9	1.2	0.6
Strong concern about speaking AND listening %	0.5	0.8	1.2

1.2.7 Educational background

The main purpose of this data set is not to examine the influences of education of carers on language development. However there may be patterns of reported language development related to education background within the sample. Analysis found that the word count scores of the

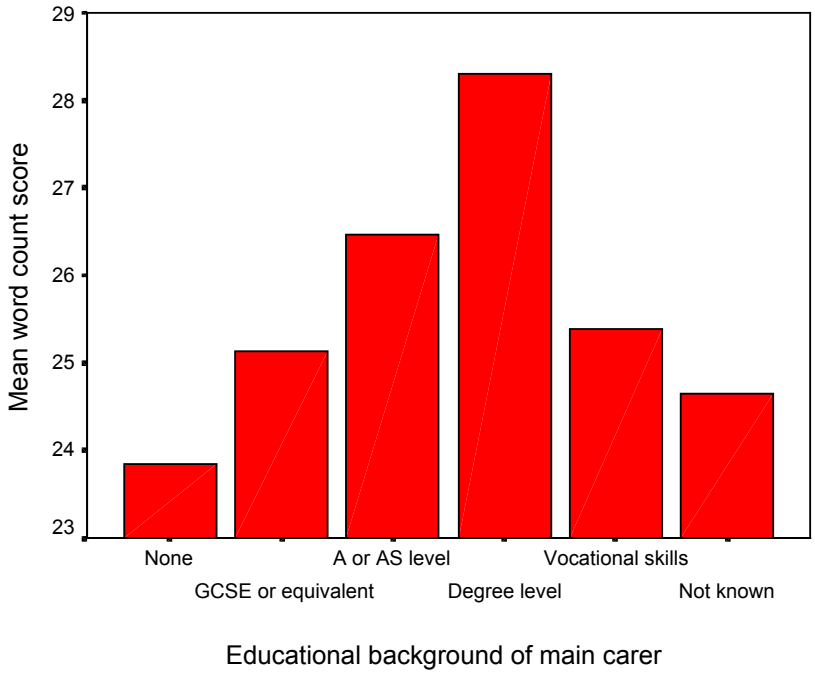
children varied with the educational background of the main carer. (F=12.615, df=5, 4688, p<0.01). Children of parents educated to degree level have the highest mean word count scores.

The lowest word count scores are from children whose carers are without qualifications or are of unknown educational level (see Table 15 and Figure 5).

Table 15: Mean word count scores by educational background of main carer for 2004 cohort

Educational background of main carer	Number of children	Percentage	**Mean word count score** (and sd)
No qualifications	920	19.6%	23.9 (12.4)
GCSE or equivalent	1743	37.1%	25.1 (12.4)
A or AS level	398	8.5%	26.5 (12.5)
Degree level	826	17.6%	28.3 (12.5)
Vocational skills	392	8.4%	25.4 (12.9)
Not known	415	8.8%	24.7 (13.2)

Figure 5: Mean word count score by educational background of main carer; 2004 cohort



The pattern of reported word combining also varied significantly across the educational groups. The parents with no qualifications, or of unknown educational background reported the lowest levels of word combining. The highest levels of word combining were reported by parents with A ,

AS and degree level backgrounds. The degree of differences across the educational backgrounds reached significance (F=8.373, df 3,4681, p=0.00). See Table 16.

Table 16: Word combining by educational background; 2004 cohort

Educational background of main carer	Percentage not combining words	Percentage combining words sometimes	Percentage combining words often
None	18.2	29.3	52.5
GCSE or equivalent	16.8	25.5	57.6
A or AS level	13.8	25.4	60.8
Degree level	10.9	22.5	66.6
Vocational skills	16.9	22.8	60.4
Not known	17.8	28.2	54.0

The parents with A, AS or degree level education were less likely to report concerns about their child's language development. The highest levels of concern were among parents of unknown educational background and those with vocational skills. The degree of variation between educational backgrounds was significant (F=3.44, df 5, 4687, p=0.004). The percentages of parents with language concerns for the different educational backgrounds are displayed in Table 17.

Table 17: Percentage of parents with language concerns, by educational background; 2004 cohort

Educational background of main carer	Percentage language concerns (%)	Number of children
None	18.8	920
GCSE or equivalent	18.8	1742
A or AS level	14.8	398
Degree level	13.9	826
Vocational skills	20.4	392
Not known	21.0	415

Boys and girls were spread evenly across the groups by educational background (See Table 18). Boys with well-educated parents have a higher mean word count score than boys of poorly educated parents. However the advantage to girls persists across the different educational backgrounds, with the boys from the well-educated backgrounds having the same mean word count as those girls whose parents have no educational qualifications (See Table 19 and Figure 6.)

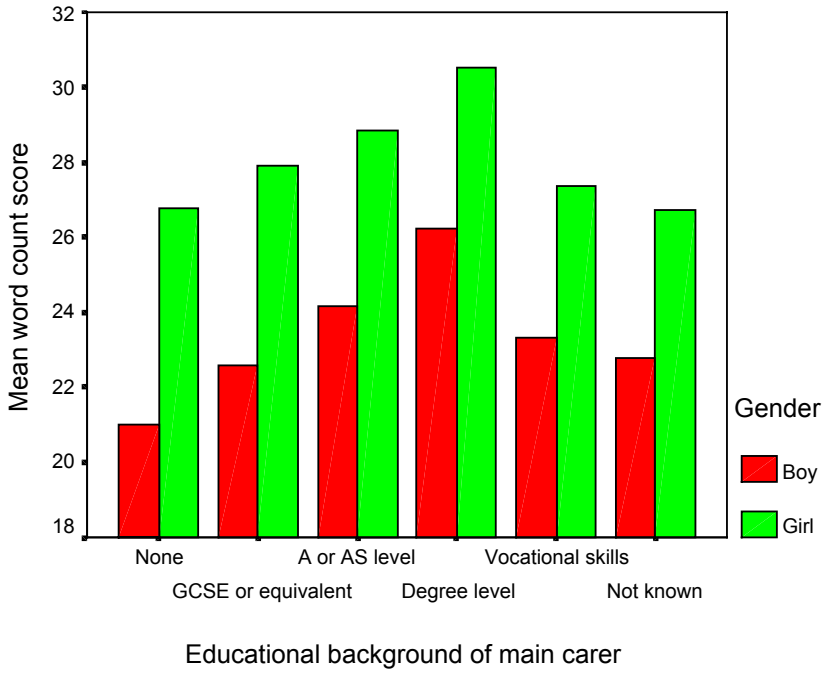
Table 18: Gender balance across the groups by educational background; 2004 cohort

Educational background of main carer	BOYS	GIRLS
None	19.3%	20.0%
GCSE or equivalent	37.6%	36.7%
A or AS level	8.4%	8.6%
Degree level	17.5%	17.6%
Vocational skills	8.1%	8.6%
Not known	9.1%	8.6%

Table 19: Mean word count scores by gender and educational background; 2004 cohort

Educational background of main carer	BOYS Mean word count score	GIRLS Mean word count score
None	21.0	26.8
GCSE or equivalent	22.6	27.9
A or AS level	24.2	28.9
Degree level	26.2	30.5
Vocational skills	23.3	27.4
Not known	22.8	26.7

Figure 6: Mean word count scores by gender and educational background; 2004 cohort



For each educational background, the boys had significantly lower word count scores than the girls. The boys also had significantly less word combining, except for the vocational skills background, and significantly higher levels of parental concern, except for the vocational and A/AS level skills backgrounds.

An analysis also looked at the educational backgrounds across the different programmes, when grouped by Round number. There is broadly the same spread of educational backgrounds for each Round, with no significant differences found ($F=1.344$, df 5, 4688, $p=0.243$). This suggests that regardless of whether the programme has been running five years or one, there is the same mix of backgrounds in the children sampled for the SSLM in the third implementation. The profile of educational backgrounds is given for each Round of programmes in Table 20.

Table 20: Educational backgrounds by programme Round number: 2004 cohort

Educational background of main carer (percentages)	ROUND 1	ROUND 2	ROUND 3	ROUND 4	ROUND 5	ROUND 6
None %	15.3	21.1	18.4	20.8	20.9	17.4
GCSE or equivalent %	37.5	36.6	37.8	37.6	36.5	38.9
A or AS level %	8.5	7.3	8.7	7.5	9.5	6.9
Degree level %	16.9	17.1	18.7	16.1	17.7	21.9
Vocational skills %	9.4	7.3	8.4	10.3	8.2	4.0
Not known %	12.3	10.7	8.1	7.6	7.2	10.9

1.2.8 Round number

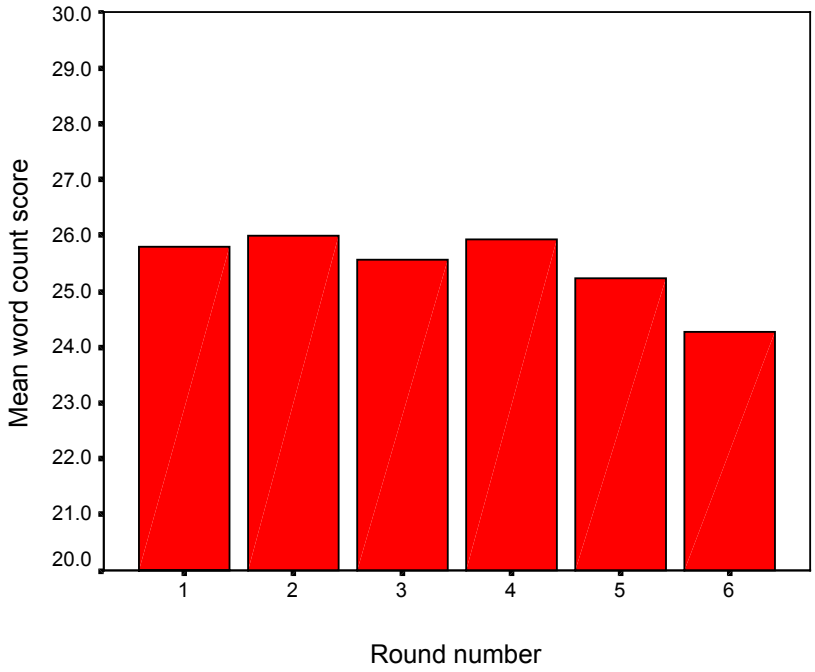
The data can be considered according to how long the programme has been operating. Programmes established earlier may report more impact on mean word count scores. The analysis looked at both individual Rounds and Rounds in subgroups of 'early' and 'late'. The differences are small.

The mean word count scores by programme Round number show only a small range of variation. There is no significant difference between Rounds when they are compared one to the other ($F=1.075$, df 5, 4688, $p=0.372$). See Table 21 and Figure 7.

Table 21: Mean word count scores by programme Round number; 2004 cohort

Round number	Mean word count score	Number of children	Std. Deviation
1	25.80	658	12.24
2	25.98	741	12.49
3	25.58	643	12.44
4	25.91	696	12.56
5	25.23	1709	12.80
6	24.26	247	13.08
Total	25.53	4694	12.60

Figure 7: Mean word count scores by programme Round number; 2004 cohort



The data set was then split into subgroups, according to Round number, to see if there was any effect for early and late established Rounds. The samples in these subgroups are similar in terms of the percentage of boys, bilinguals, education level of carer and special needs. They do however vary in terms of their sampling techniques[9] (see Table 22).

[9] Early / late rounds by sample source: F=9.667, df 1, 4692, p<0.01.

Table 22: Features of early and late programme subgroups by programme Round number

	EARLY PROGRAMMES Rounds 1-4	LATE PROGRAMMES Rounds 5-6
Boys (%)	51.3	51.8
Language background (%)		
English	74.6	74.2
Bilingual	21.5	22.6
Non-English	3.8	3.1
Special needs children %	2.3	2.3
Education level of carer (%)		
No qualifications	19.0	20.4
GCSE	37.4	36.8
A/AS level	8.0	9.2
Degree level	17.2	18.2
Vocational skills	8.8	7.7
Unknown	9.7	7.7
Sample source (%)		
Representative	34.6	38.2
Sure Start	32.3	33.2
Mixed	18.2	19.8
Unknown	14.9	8.7

Table 23 shows only a modest contrast in the word count means between Early programmes (Rounds 1 to 4) compared to Late programmes (Rounds 5 and 6). Early programmes had a slightly higher word count mean of 25.8, compared to Late programmes with a mean of 25.1. (This contrast did not reach statistical significance, as indicated by an ANOVA comparison: $F=3.712$, df 1, 4692, $p=0.054$). Likewise the difference in levels of word combining failed to reach significance ($F=1.13$ df 1, 4685, $p=0.288$). The levels of parental concerns about language and general development are however fairly stable regardless of the programme's duration. The slightly lower mean word count for Late programmes is reflected in the percentage S being slightly reduced compared to Early programmes.

Table 23: Key results by Early and Late programmes; 2004 cohort

2004 FULL DATA SET	EARLY programmes (N=2738) Rounds 1 to 4	LATE programmes (n= 1956) Rounds 5 and 6
Mean word count score (sd)	25.8 (12.4)	25. 1 (12.8)
Children combining words Sometimes or often (%)	84.7	83.2
Parental language concerns (%)	18.2	17.5
S %	74.8	72.9

The cross-tabulation of word count score and parental language concern for the Early and Late Rounds in Tables 24 and 25 shows how the percentage S is raised for the Early programmes. The difference in S between Early and Late-established Rounds is significant, using a chi-square test (chi-square=30.18, df 3, $p<0.05$). This is the net effect of a marginally higher word count mean for Early programmes, even though the parental language concern is slightly higher in the earlier programmes. The effect could be due to sample source variation between the two subgroups. The data does not allow a clear conclusion about the cause of the raised percentage S for Early programmes. It cannot be assumed to be due to programme impact, because the influence of sample (demographic) variables cannot be ruled out.

Table 24: Word count scores and parental language concerns for 2004 cohort: Early Rounds 1-4

	Word count <=12	Word count >=13	TOTAL
Parental language concern YES	255	243	498
Parental language concern NO	193	2047 S=2047 / 2738 = 74.8%	2240
TOTAL	448	2290	2738

Table 25: Word count scores and parental language concerns for 2004 cohort: Late Rounds 5-6

	Word count <=12	Word count >=13	TOTAL
Parental language concern YES	193	150	343
Parental language concern NO	186	1426 S= 1426 / 1955 = 72.9 %	1612
TOTAL	379	1576	1955

1.3 Summary

The third implementation data is presented in terms of key findings. These are then explored for relationships between groups by gender, language background, educational background of main carer, and by the Round number of the Sure Start local programme.

The language background of the child, the programme Round number and the sample source (how the children were selected) did not produce an effect on the word count scores. Word count scores varied for boys and girls, with girls scoring 28 out of 50 to the boys mean score of 23. The educational background of the main carer also produced an effect on word count scores, with the more highly educated carers reporting higher word count scores for their child. This effect held for both girls and boys.

74% of the parents reported no language concerns and a high word count score for their child.

Part 2 Comparison over three implementations

2.1 Presentation of results

As for the third implementation, the comparative analysis gives key results for word count score, word combining, and the level of parental concerns about language development. These results are again summarised into percentages that combine word count score with parental language concerns. Patterns of data are also presented by gender and by language background.

2.2 The comparison data set

In order to look at change over time, a selection of data from each of the three years' datasets was compared. The word count scores were all compared out of 50. The procedure for selecting a proportion of each year's data tried to avoid selection bias. The population characteristics of these subsets were in many respects similar to the overall data set for each implementation. There was similarity for sample types[10], region, gender, special needs, age of child, and maternal indicators (age of mother at first birth and educational background). 38% of the comparison dataset was sampled in a representative way, with 25% sampled from Sure Start registrations.

There is however a difference in the proportion of language backgrounds in the comparison dataset. 81% of the comparison dataset consisted of English speaking families, compared to 75% in the remaining data not used for the comparison dataset. Even though there were relatively fewer bilinguals than would be expected in the comparison dataset, their average word count scores were higher than those of the bilinguals not used for the comparison dataset. This introduces the possibility of bias in the comparison dataset for the results by language background. Consequently robust conclusions over the three years can be drawn only for the English-speaking language category. The construction of the comparison set and its population characteristics are set out in full in Appendix 3.

2.3 Comparisons over three implementations

2.3.1 Key results

The mean word count has risen from 25.3 to 26.2 out of 50 over the three implementations (see Table 26). There is not a significant level of change from either 2001 to 2004 or 2003 to 2004. [2001 to 2004: F= 1.66, df 2,3632, p=0.191. 2003 to 2004: F= 1.195, df 1,2709, p=0.274.] Within this observation, the cases sampled only from Sure Start programmes show more change than those cases sampled in a representative or wide sampling manner. For Sure Start only samples, the mean word count rose from 25.3 (2001) to 26.2 (2003) to 26.8 (2004). However neither do these changes for the Sure Start only subgroup reach statistical significance over the three implementations (F=0.577, df 2, 905, p=0.562).

[10] Representative: Children's names selected in an unbiased way from a comprehensive listing of two year olds
Wide sample: Children's names selected from various sources, including opportunistic contacts
Sure Start only: children's names taken from local programme registrations
Mixed approaches: combination of the above
Unknown: sample source not specified

Table 26: Key results compared over three implementations: all language backgrounds

IMPLEMENTATION YEAR	**2004**	2003	2001
Number of children	**1409**	1302	924
Mean word count score (sd)	**26.2** **(12.5)**	25.7 (12.8)	25.3 (13.0)
Children combining words Sometimes or often (%)	**84.6**	84.9	84.7
Parents with language concerns (%)	**17.9**	19.3	21.9

The drop in levels of parental language concern reaches significance over the three years, from 22% (in 2001) to 18% (in 2004) (Kruskal-Wallis chi-square = 5.63, df 1, 2333, p=0.018).

For English-speaking children only, there is the same pattern (see Table 27). The mean word count score is rising over the three implementations, but without reaching a statistically significant level (F=1.197, df 2, 2940, p=0.302). There is no significant influence from sampling techniques on the English-only group of cases. That is, within each implementation, the mean scores do not vary significantly depending on the sampling technique. Also, across the three implementations, there is no significant change in the mean word count score for each sample source taken separately.

Table 27: Key results compared over three implementations: English-speaking children only

IMPLEMENTATION YEAR	**2004**	2003	2001
Number of children	**1097**	1046	800
Mean word count score (sd)	**26.3** **(12.5)**	26.0 (12.7)	25.4 (12.9)
Children combining words Sometimes or often (%)	**85.0**	87.1	85.7
Parents with language concerns (%)	**17.9**	18.9	21.9

2.3.2 Distribution of word count scores

As well as looking at the mean word count score, it is appropriate to look at the distribution of the scores. For each implementation, the distribution of scores can be plotted on a graph, showing the spread of low and high scorers. In 2001, the word count scores showed a left-hand lean (positive left skew). See Figure 8. This suggested a number of possibilities for change over time. It was hypothesised that the distribution might even out to a smoother shape, showing relatively more improvement in the lower scores than in the higher scores. This would be recognised by an unchanged mean, but different skew indicators. Alternatively, the whole curve might shift to the right, showing that all scores had improved over time, but retaining a left-hand leaning shape. In this scenario the mean score would rise but the shape of the curve would be unchanged. A final possibility was that there would be a combination of such effects, showing both a rise in the mean score and a change in the distribution skew indicators.

Accordingly the distribution indicators for the comparison data set are given below. (See Tables 28 and 29). There is the same pattern for the full comparison data set as for the subset that is English-speaking only. The skew indicator moves from positive to negative, showing that the curve is leaning more to the right than to the left. This suggests that the first scenario of change may be operating. There is relatively more improvement in the lower scoring children than in the higher scoring children. This change is mild in 2003 and increases in 2004. See Figures 9 and 10.

Another aspect of the distribution of word count scores is to look at the sample types separately. (See Table 30.) The centile scores for the Sure Start only children show that there is a small rise in word count scores at each centile for all the scores except the lowest 14 percent. This observation captures the fact that children sampled from Sure Start registrations are showing a small descriptive rise (between 2001 and 2004) in their scores across the ability spectrum, except for the lowest 14 percent. However, for children sampled in a representative way, the scores above the mean are moving down toward the mean, while scores below the mean are stable.

There is the possibility that the distribution changes over the three years are due to variation in the population descriptors. For example, the third implementation could have more well-educated parents than earlier cohorts. Unfortunately the education variable is not available in all three years for comparisons. Thus the proportion of distribution change due to sample variables cannot be estimated.

Table 28: Distribution indicators over the three implementations; comparison subset

IMPLEMENTATION YEAR	**2004**	2003	2001
Number of children	**1409**	1302	924
Mean word count score (sd)	**26.2 (12.5)**	25.7 (12.8)	25.3 (13.0)
Standardised skew statistic	**-1.23**	-0.09	+0.84
Distribution lean	**right**	right	left

Table 29: Distribution indicators over the three implementations; English-only cases

IMPLEMENTATION YEAR	**2004**	2003	2001
Number of children	**1097**	1046	800
Mean word count score (sd)	**26.3 (12.5)**	26.0 (12.7)	25.4 (12.9)
Standardised skew statistic	**-0.73**	-0.05	+1.0
Distribution lean	**right**	right	left

Table 30: Word count scores for the centiles in 2001 and 2004 cohorts

CENTILE	ALL SAMPLES		Representative samples only		Sure Start samples only	
	2004	2001	**2004**	2001	**2004**	2001
90	**43**	43	**42**	44	**44**	43
80	**38**	38	**37**	39	**39**	36.4
75	**36**	35.75	**35**	36	**36.5**	34
70	**33**	33	**33**	34	**35**	32.6
60	**30**	29	**29**	30	**31**	30.8
50	**27**	25	**26**	25	**27**	25
40	**23**	22	**21**	22	**24**	21
30	**19**	17	**17**	17.2	**20**	16
25	**17**	15	**15**	15	**17**	15
20	**15**	13	**13**	13.8	**15**	14
15	**12**	11	**10**	12	**12.9**	11.2
10	**8**	7	**8**	8.4	**8**	8.8
5	**5**	4	**5**	5	**5**	7

Figure 8: Distribution of word count scores in the comparison data set; 2001 cohort

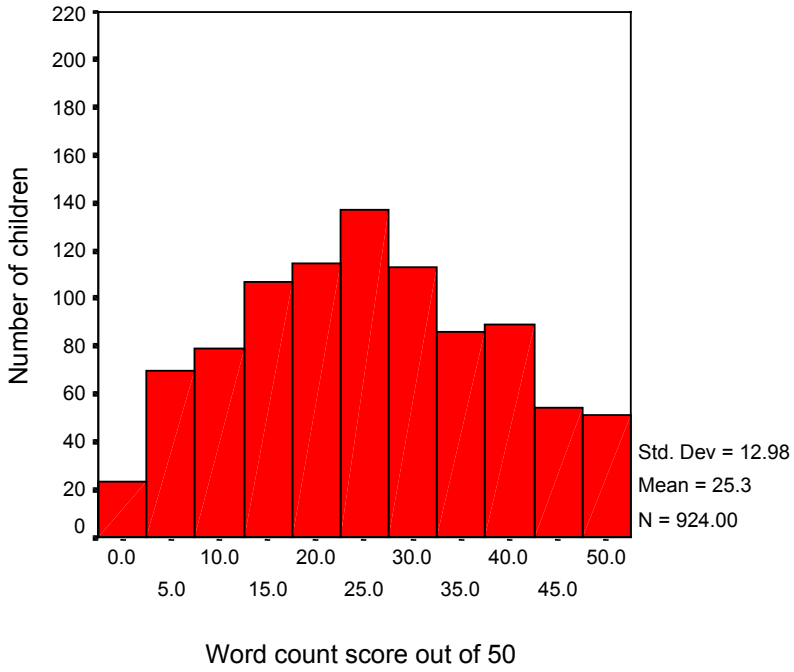


Figure 9: Distribution of word count scores in the comparison data set; 2003 cohort

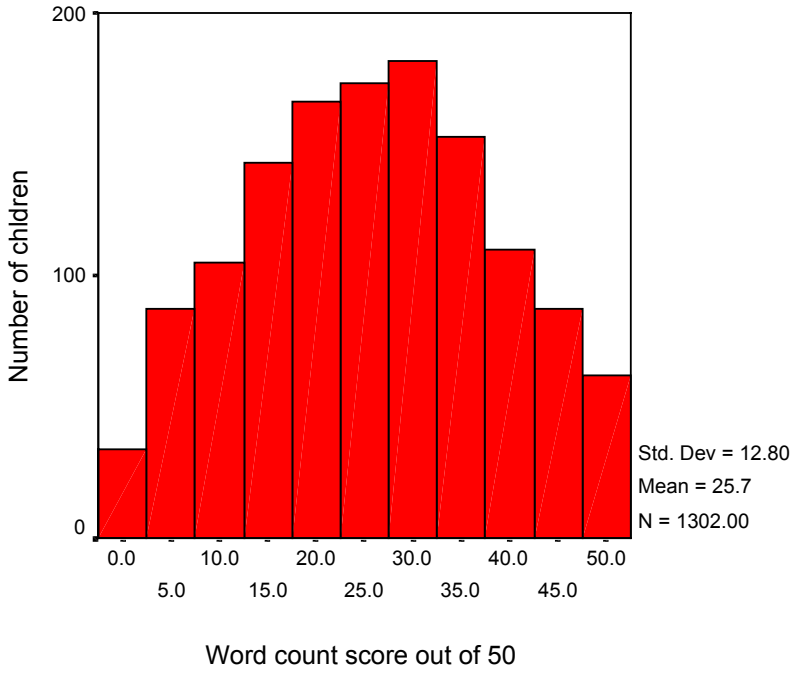
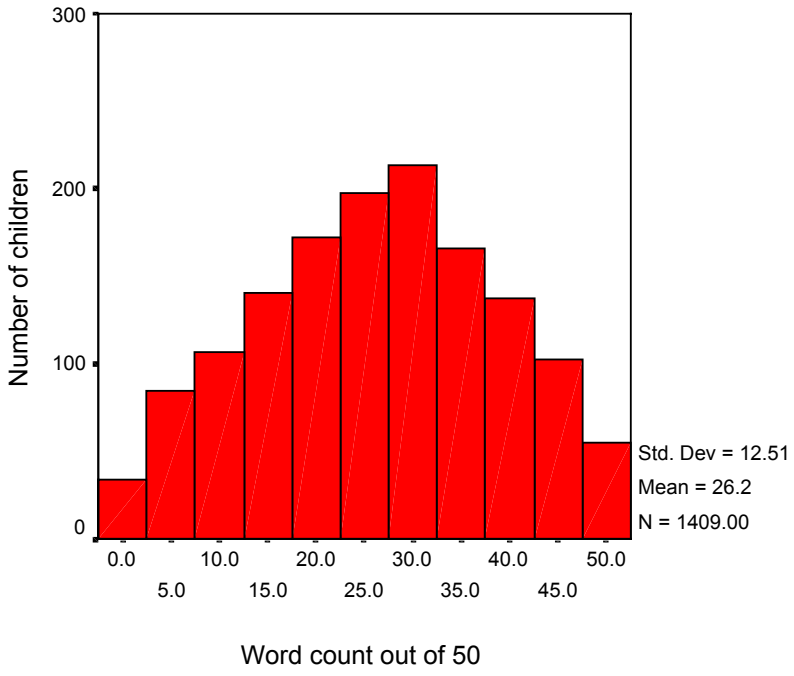


Figure 10: Distribution of word count scores in the comparison data set; 2004 cohort



2.3.3 Parental report of concerns

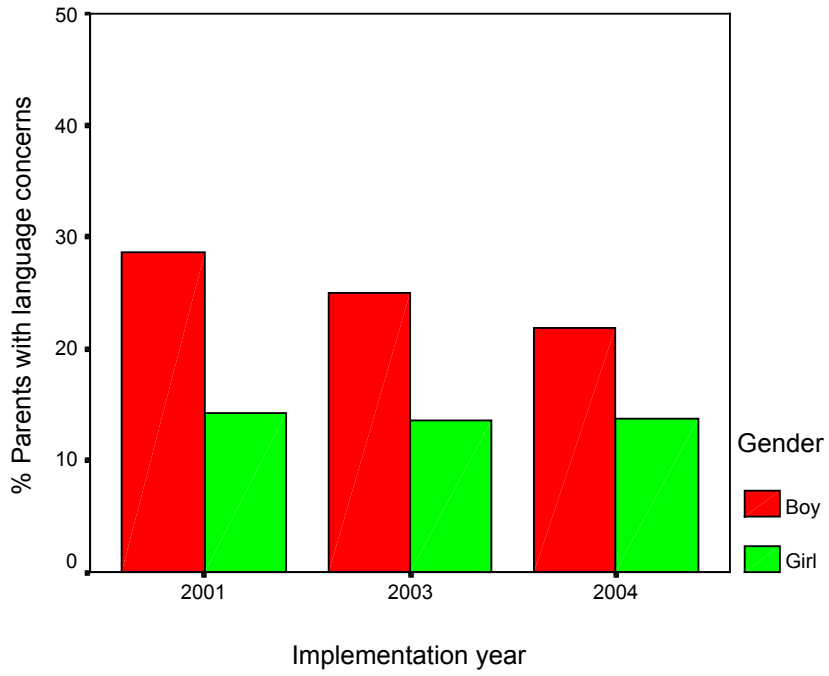
Parental concerns about general and language development were recorded using the PEDS[11] questionnaire (in 2001 and 2003) and using a set of four questions (in 2004). The questions about language development were broadly comparable, and show a drop in the percentage of parents with concerns about their child's language development (see Table 31). This is mainly attributable to parents of boys becoming less concerned (see Figure 11).

Table 31: Parental language concerns over the three implementations

IMPLEMENTATION YEAR	**2004**	2003	2001
Number of children	1409	1301	924
Parents with language concerns	17.9 %	19.3 %	21.9 %

[11] Parental Evaluation of Developmental Status PEDS; Frances Glascoe 1997

Figure 11: Parental language concerns for boys and girls over the three implementations



The reduction in levels of language concern across the word count scores is shown in Table 32, where children's scores are divided into low, low-average, above-average and high. The percentage shown is the proportion of children in that score grouping whose parents are concerned about their language.

This suggests that the levels of language concern are falling across the low-average and high ranges of word count scores, rising for the low scorers, and fairly stable for the above-average score group. The most marked drop in concern is for the low-average group, being the children who are showing the most improvement in their scores between 2001 and 2004.

Table 32: Parental language concerns by score group over the three implementations

Score level	**2004**	2003	2001
LOW Score 12 or less	**60.2 %**	55.3 %	58.1 %
LOW-AVERAGE Score 13 to 26	**15.8 %**	17.9 %	21.4 %
ABOVE-AVERAGE Score 27 to 36	**8.1 %**	8.5 %	7.9 %
HIGH Score 37 to 50	**3.1%**	6.0 %	6.2 %

The questions on general development show less continuity between 2001/2003 and 2004, probably because of the change in the questions used. 93% of parents reported no concerns about general development in 2004. This figure probably excludes concerns about language development. However this figure does not compare with earlier percentages of 57% and 63% and it is not easy to interpret. Time did not allow examination of response forms to discover patterns in responses. Table 33 shows general developmental concerns over the three implementations.

Table 33: General developmental concerns over the three implementations

General developmental concerns %	**2004**	2003	2001
None[12]	**92.6**	63.0	57.0
Some concerns[13]	**5.6**	29.9	37.2
Strong concerns[14]	**1.9**	7.1	5.8

2.3.4 Summary figures

The summary figures are considered as indicators of change over the three implementations. The percentages P and Q represent children about whom there would be concern, due to parental concerns and/or a low word count. Thus over time reduced P and Q percentages would be desirable. A rising percentage S would represent more children free of parental language concern and with the higher word counts. The percentage R might be stable over time, or reducing: this would represent children for whom a low word count and no parental concern is appropriate, such as low-performing children where the parents have developmentally appropriate expectations.

For each implementation, Tables 34-36 set out the numbers of children in each category, according to their word count score and the presence or absence of parental language concern. From these the percentages PQ R and S are presented in Table 37. (See Table 8 for methods of calculating P Q R and S.)

[12] 2001 and 2003; PEDS category E: no concerns
[13] 2001 and 2003; PEDS categories B and C: some minor concerns or one major concern
[14] 2001 and 2003; PEDS category A: two or more major concerns

Table 34: Word count scores and parental language concerns for comparison data set (2004)

	WORD COUNT SCORE 12 or less	WORD COUNT SCORE 13 or more	TOTAL
PARENTAL LANGUAGE CONCERN			
Yes	136	116	252
No	90	1067	1157
TOTAL	226	1183	1409

Table 35: Word count scores and parental language concerns for comparison data set (2003)

	WORD COUNT SCORE 12 or less	WORD COUNT SCORE 13 or more	TOTAL
PARENTAL LANGUAGE CONCERN			
Yes	125	126	251
No	101	949	1050
TOTAL	226	1075	1301

Table 36: Word count scores and parental language concerns for comparison data set (2001)

	WORD COUNT SCORE 12 or less	WORD COUNT SCORE 13 or more	TOTAL
PARENTAL LANGUAGE CONCERN			
Yes	100	102	202
No	72	650	722
TOTAL	172	752	924

Table 37 shows how the percentage S has risen over the three implementations, using the comparison subset of data. This is the proportion of children scoring 13 or more out of 50 on the word list, and with parents free of concerns about their child's language. For the purposes of comparison, PQR and S for the first and second implementations are calculated on scores out of 50. S rises because of the combined impact of marginally higher word count scores, and significantly reducing levels of parental language concern.

Table 37: Summary figures PQRS over the three implementations (comparison data set)

All figures percentages	**2004**	2003	2001
P	**9.7**	9.6	10.8
Q	**8.2**	9.7	11.0
R	**6.4**	7.8	7.8
S	**75.7**	72.9	70.3
Total number of children	**1409**	1301	924

With the rise in percentage S, there is a fall in the percentages PQ and R over the three implementations. There are fewer children whose parents are concerned about their language (relative to the size of the sample), and fewer children with low word counts and parents without language concern. The change between the 2001 baseline and 2004 reaches statistical significance (chi-square =14.02, df=3, $p<0.05$); this is not the case for the change between 2003 and 2004.

A more conservative approach considers only the monolingual English speakers, leaving out the bilingual groups from the comparison subset, due to possible selection bias. Table 38 gives the PQRS figures for English-only language backgrounds over the three implementations. The change between the 2001 baseline and 2004 reaches significance (chi-square=11.17, df=3, $p<0.05$.).

Table 38: Summary figures PQRS over the three implementations (comparison data set; English -only)

All figures percentages	**2004**	2003	2001
P	**9.8**	9.3	10.8
Q	**8.1**	9.7	11.1
R	**6.2**	7.3	7.3
S	**75.9**	73.8	70.9
Total number of children	**1097**	1046	800

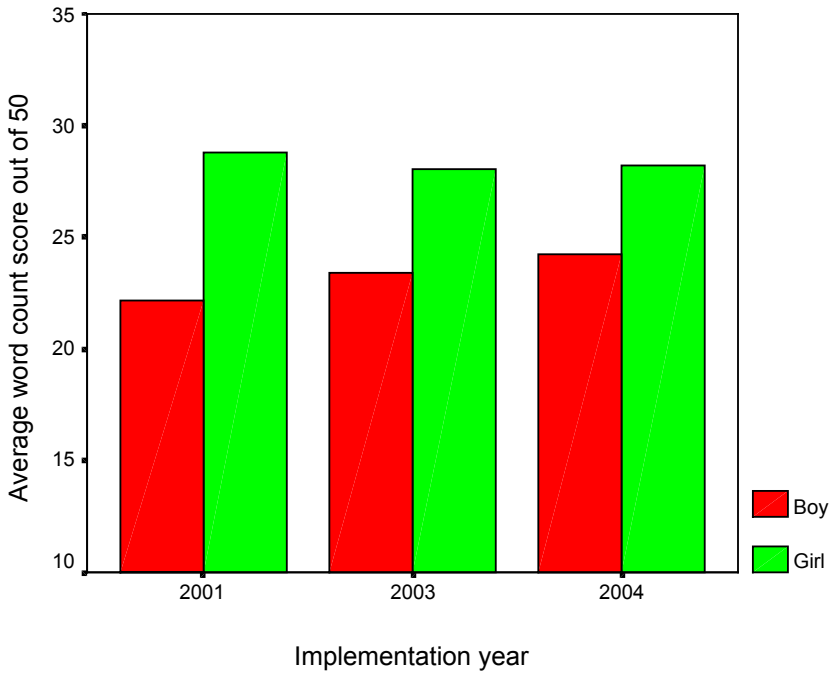
2.3.5 Gender

The mean word count scores for boys and girls over the three implementations are given in Table 39. Over the three implementations, the girls maintained a higher word count score than the boys. (In each implementation the difference between the girls and boys reached statistical significance.) Of interest is the boys' mean scores rising over the three implementations, while the girls' score is unchanged. Considering the boys only, there is a statistical difference in their scores over the three implementations ($F=3.858$, df 2, 1859, $p=0.021$). Statistical significance is not achieved when comparing the boys from 2003 to 2004 ($F= 1.53$, df 1,1371,$p=0.216$).

Table 39: Mean word count score for boys and girls over the three implementations

IMPLEMENTATION YEAR	**2004**	2003	2001
BOYS Number of children N=1862	**716**	657	489
Mean word count score (Sd)	**24.3 (12.9)**	23.4 (12.7)	22.2 (12.4)
GIRLS Number of children N= 1764	**687**	644	433
Mean word count score (Sd)	**28.2 (11.7)**	28.1 (12.3)	28.8 (12.8)

Figure 12: Mean word count scores for boys and girls over the three implementations



The distribution of word count scores shows different patterns for boys and girls. Considering the change between 2001 and 2004, the boys' scores have all improved across the range of scores from 0 to 50. This means that both low and high ability boys were able to improve their scores between 2001 and 2004.The mean rises and the whole distribution shifts to the right. For girls, the mean remains the same, and scores are less tightly spread around the mean. In other words the lowest scores improve and the highest scores reduce. This suggests that for girls of all abilities their scores are moving towards the mean score.

The rise in word count scores holds when all the boys of different language backgrounds are taken together. When considering only the English-speaking boys, there is still a descriptive improvement in their mean scores, but this does not reach statistical significance. (F=2.771, df 2, 1492, p=0.063). The mean word count scores by gender for the English-speaking children are shown in Table 40.

Table 40: Mean word count score for English-speaking boys and girls over the three implementations

IMPLEMENTATION YEAR	**2004**	2003	2001
BOYS Mean word count scores (sd)	24.1 (12.7)	23.7 (12.7)	22.3 (12.3)
GIRLS Mean word count score (sd)	28.3 (12.0)	28.3 (12.3)	29.0 (12.7)

There is no clear trend in the rates for word combining over the three implementations for the boys and girls (see Table 41). This is also the case for the English only subset.

Table 41: Word combining by gender over the three implementations; comparison data set

IMPLEMENTATION YEAR	**2004**	2003	2001
BOYS Not combining words	**19.4**	21.2	20.0
Sometimes	**27.3**	23.5	28.0
Often	**53.3**	55.3	51.9
GIRLS Not combining words	**11.5**	9.0	9.9
Sometimes	**21.4**	19.5	19.2
Often	**67.1**	71.5	70.9

For boys, parental language concern has decreased by nearly 7 per cent over the three implementations, while for girls it has reduced by less than one per cent (see Table 42). The change in parental language concern for boys is significant when comparing 2001 to 2004, but not for 2003 to 2004. [2001 to 2004: F=3.5, df 2, 1858, p=0.03; 2003 to 2004: F= 1.8 df 1, 1370, p=0.18.]

Table 42: Parental language concern by gender: comparison data set

IMPLEMENTATION YEAR	2004	2003	2001
BOYS N=1861	**21.9%**	25.0 %	28.6%
GIRLS N= 1764	**13.7%**	13.5%	14.3%

2.3.6 Language background

The word count scores out of 50 were compared over the three years for the language background groups: English-speaking only, bilingual with English, and non-English speaking. In each implementation, the word count scores did not differ by language background. (The first and second implementations showed the 100 word count scores differing by language background.) There are two possible reasons for this, which are not mutually exclusive. It could be that the bilinguals in the comparison set are the more able bilingual cases, with higher mean scores. Alternatively the 50 word list could more fairly report the language skills of children from bilingual backgrounds. It is likely that both influences are operating. (The results from the full third implementation cohort suggest that the 50 word list is at least an improvement for the bilingual groups.)

For English speaking children, scores rose from 25.4 (2001) to 26.3 (2004), while for children using English and another language, their mean word count score rose from 24.8 (2001) to 26.4 (2004). Scores for non-English speaking children rose from 23.8 (2001) to 24.8 (2004). See Table 43 and Figure 13.

The scores have to be considered carefully in view of the possible selection bias for this subset of recounted cases. The bilinguals in the comparison subset may be the more able ones, unintentionally boosting their scores out of 50. For more detail regarding the bilinguals and English-only groups in the comparison dataset please see Appendix 3.

Figure 13: Mean word count scores by language background, over the three implementations

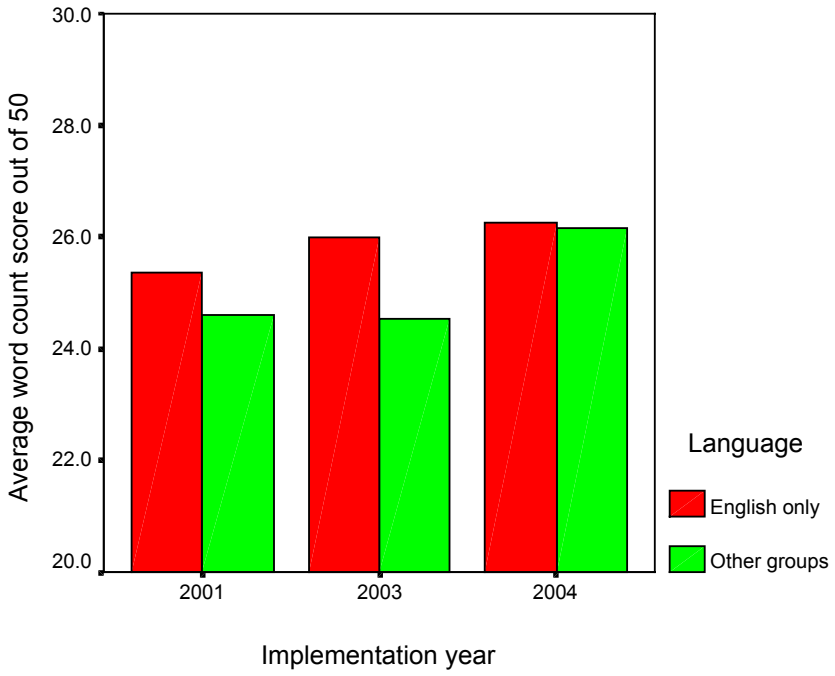


Table 43: Mean word count scores by language background across the three implementations

IMPLEMENTATION YEAR	**2004**	2003	2001
ENGLISH SPEAKING ONLY N= 2943 Mean word count score (sd)	26.3 (12.5)	26.0 (12.7)	25.4 (12.9)
BILINGUAL WITH ENGLISH N= 565 Mean word count score (sd)	26.4 (12.3)	24.2 (12.7)	24.8 (13.2)
NON-ENGLISH SPEAKING N= 125 Mean word count score (sd)	24.8 (13.9)	25.8 (15.5)	23.8 (14.3)

With regard to the distribution of word count scores, the English-speaking children showed a small rise in the word count score for the 10 to the 70th centiles, with the tails of the distribution unchanged. Thus mid-ability English children were benefiting most. For children bilingual with English, the scores below the mean showed a small improvement while the scores above the mean were stable, showing that the low-average and lower-scoring bilingual children were showing most change from 2001 to 2004.

2.3.6.1 Word combining

The overall levels of word combining have remained stable over the three years. Looking at the subgroups by family language background, the levels of word combining have been rising for bilingual and non-English speaking groups, but not for English-speaking children. The smaller groups sizes of the bilingual and non-English groups mean that the overall levels of word combining do not show change over the three years. See Table 44.

Table 44: Levels of word combining by language background; comparison data set

IMPLEMENTATION YEAR	**2004 % combining words sometimes or often**	2003 % combining words sometimes or often	2001 % combining words sometimes or often
ENGLISH SPEAKING ONLY N= 2943	**85.0**	87.1	85.7
BILINGUAL WITH ENGLISH N= 565	**84.3**	77.2	79.6
NON-ENGLISH SPEAKING N= 125	**76.5**	70.0	70.8

2.3.6.2 Parental language concerns

For the full comparison data set, the trend is for falling percentages of parental language concern. The data by language background show that for non-English family backgrounds the language concerns have been rising. The trend may be due to variation within the small sample sizes. The very small size of this group means that its contribution to the full set is marginal. See Table 45.

Table 45: Percentage of parents with language concerns, by language background; comparison data set

IMPLEMENTATION YEAR	**2004**	2003	2001
ENGLISH SPEAKING ONLY N= 2943	17.9	18.9	21.9
BILINGUAL WITH ENGLISH N= 565	16.9	20.4	21.4
NON-ENGLISH SPEAKING N= 125	23.5 (n=51)	22.0 (n=50)	20.8 (n=24)

2.4 Summary

The results over three years are compared using a selected subset of the data. Key findings show a very small rise in the mean word count score, a drop in parental language concerns and no change in the rate of children combining words.

Boys have narrowed the gap in word count scores to the girls, but remain behind both in terms of their mean word count score and their tendency to have a lower spread of scores.

Language background of the family does not influence the word count score, based on the comparison subset.

Analysis of the distribution suggests that the children showing most change in their scores between 2001 and 2004 are the low-average children, scoring between the 15th and 50th centiles. Their scores have risen and the reported levels of parental concern have dropped. These changes in turn have raised the percentage level S, which indicates the percentage of parents with no language concerns and reporting high word count scores for their child. Over the three years S has risen from 70% to 76%, based on the comparison subset.

2.5 Discussion

2.5.1 The comparison data set

The analysis over three implementations relies on the comparison data set, formed by taking a subset of each year's data. The approach used for selecting this set was intended to be free of bias. The only assumption made was that by including the Early programmes that had sent data repeatedly, there would be more consistency of demographic indicators, and more likelihood of detecting changes. A detailed analysis of the comparison set showed that, unintentionally, the bilingual component (those other than English-speaking) was under-represented in terms of numbers but these bilinguals were relatively more able. As a result the analysis has tried to

distinguish wherever possible between the English-only part of the comparison data set and the full set. This has showed that the results hold for the English-only subset, where we can be most confident that the subset reflects the full cohort over the three years. A more balanced set for comparison could be achieved by further recoding of the first and second implementations bilingual cases to obtain their scores out of 50. (Project time constraints have not allowed this.)

It remains unclear why a selection approach favouring Early programmes should have sampled the more able bilingual cases. It suggests that Early-established programmes had more able bilingual participants. This could be a start-up effect where well-educated bilingual families take up Sure Start services earlier than other bilingual families.

2.5.2 Boys and girls

A clear result for the three implementations is the narrowing gap between scores of the boys and girls. The girls have maintained their advantage over boys by a significant margin each year. The boys however have improved their word count scores while the girls have maintained their level. It is not clear why it is only the boys who have improved their scores. The rise in scores could be related to the family background (particularly the educational background of the main carer) of the boys in the third implementation. It is clear from the third implementation data that the educational background impacts on the mean word count score of each child. There is no comparable data in the first and second implementations with which to compare the family backgrounds of the children. If we could exclude background as an influence on the children's scores, then we could consider in more detail the programme impact. Is there relatively more programme impact on boys' language? Are parents of boys more likely to seek advice about language?

2.5.3 Distribution of word count scores

The analysis looked at the distribution of word count scores over the three years. The key point of interest is: to what extent is the distribution of SSLM data different to the representative UK population, and how is the distribution changing over time? The data here has tried to address the issue of change. In the three years, with a near static mean, the word count scores have moved from showing a left hand skew towards a small degree of right hand skew. This is positive in as much as the low-average children, in the 15 to 50th centile range, are improving their scores. (The scores above 60th centile are more stable). There remains a persistent tail of low-scoring children below the 15th centile. As far as comparison to UK norms are concerned, a separate study for standardising the SSLM across a range of representative UK 18-30 months olds is underway and should report during 2005 (Law et al, forthcoming).

2.5.4 Parental concern about language development

The reduction in parental concern about language development is taken as a positive change. Of itself, reductions in concern could be an indicator of parents who are not well informed or not engaged with their child's development. However here the reduction in parental language concern is shown particularly in the low-average section (15th to 50th centile) of the distribution of word count scores, where most change is taking place in word count scores. The rise in the summary percentage S captures the impact of this reduction in rates of language concern, combined with the improvement of scores in the low-average range.

2.5.5 Sample sources

Throughout the analysis attention is given to the manner of sampling the cases. Sample source does not impact to a significant level, but at a descriptive level. The margins of difference between sample sources are all small. It is hard to detect differences between these sample groups because there is overlap. A group of representatively sampled children is likely to include some Sure Start registered children. Further the other sample source categories are very varied, or unknown, and may also include a portion of Sure Start registered children. Thus the group of exclusively Sure Start children is being compared to the other groups that are probably partially made up of Sure Start registered children. The situation probably arises because the sample source types remain hard to define at the practitioner level. Another complication in interpreting sample source impact is that Sure Start registration is not equivalent to uptake of Sure Start services. Nonetheless, the Sure Start sample group shows a trend for increasing mean word count score, and for distribution indicators showing improvement for all but the lowest scoring children.

2.5.6 Power

The size of the samples over three years is impressive. The full number of cases over three years is 9177, of which 3635 have been compared in detail over the three years. There is considerable statistical power as a result. Significant findings in this set are likely to be robust. The limitation on the data set is the quality of the indicators collected, which limits the ability to draw clear conclusions.

Educational background information will prove very useful in future data collections as a way of ensuring comparability of samples. (It was not collected in the first and second implementations).

Part 3 Conclusions

3.1 Conclusions

Data from the three years leads to some general conclusions about the SSLM. The intent of the data collection was to monitor changes rather than to explore their causes. The present data cannot address issues of sample variation and changes in demographic details from year to year.

The gap between girls and boys in reported vocabulary skills has narrowed, with boys making up some of the lag in skills between 2001 and 2004. There remains an imbalance in the distribution of scores, such that boys are over-represented in the lower half of the distribution. Of interest in future datasets is the continued improvement of boys' scores, and the spread of their scores in comparison to the girls.

The revision of the SSLM to a 50 word format has been well received, as easier and quicker to use. It has also neutralised the differences that were seen between children of different language backgrounds when using the 100 word format. The analysis based on a selected subset from each year's data may be subject to selection bias. The scores of children from different language backgrounds will be compared carefully in future datasets to examine this relationship further.

The movements in the mean word count score and the distribution of scores have been carefully examined. More than the mean score, the movements in word count scores across the ability range will continue to be of interest. As Sure Start programmes continue to operate, will the scores at the very low end of the distribution show improvement? Or is this a segment of the population that is most resistant to change?

Comparison over the three implementations shows a reduction in levels of parental concern about their child's language, for parents of both low-average and high scoring children. The concerns expressed by parents about their child's language development are moving in a direction consistent with their reporting of vocabulary skills on the SSLM. This suggests that the original internal validity (consistency) of the SSLM has been maintained.

3.2 Future data collections

Further national data collections using the SSLM are planned for February 2005 and for February 2006, which would make a stream of five datasets for comparison. Programmes are also encouraged to use the SSLM for monitoring local progress towards the PSA target.

Acknowledgements

To the programmes who submitted the data
To the originators of the MCDI
To the Sure Start Speech and language steering group

Further information

Copies of the summary report can be obtained from the Sure Start website:

www.surestart.gov.uk

For further copies of the summary please contact:

DfES Publications,
PO Box 5050,
Sherwood Park,
Amesley Park, Nottingham, NG15 0DJ
Tel 0845 6022260
Fax 0845 6033360
Textphone 08456055560
Email dfes@prolog.uk.com

This technical report will also be available on the Sure Start website.

For technical enquiries about this report please contact:

James Law

Centre for Integrated Healthcare Research
Queen Mary University College
Edinburgh EH12 8TS

jlaw@qmuc.ac.uk

For other enquiries please contact:

Kiran Sidhu Rajkiran.SIDHU@dfes.gsi.gov.uk

References

Dale PS, Dionne G, Eley TC and Plomin R (2000) *Lexical and grammatical development: a behavioural genetic perspective.* Journal of Child Language 27; 619-642

Department for Education and Skills (2003) *The Level Of Highest Qualification Held By Young People And Adults: England 2002.* Nottingham: DFES

Harris F, Law J & Roy P (forthcoming) The development of a 50 word adaptation of the UK MCDI short form for use with Sure Start programmes in England.

The revised SSLM Data Collection Information Pack 2003-2004; Sure Start Unit, London, 2002.

Roy P, Kersley H and Law J (forthcoming) The Sure Start Language Measure (SSLM) standardisation study.

Appendices

1 SSLM format 2001 and 2003
2 SSLM format 2004
3 Comparison data set
4 PQRS calculations

Appendix 1: SSLM format 2001 and 2003

PROGRAMME NAME:_____

[1] CHILD IDENTIFIER _____

[2] BOY / GIRL (circle)
0 1

DATE OF SSLM _____

CHILD DATE OF BIRTH _____

[3] CHILD AGE AT SSLM _____completed months

[4] LANGUAGE Please mark one: 1 2 3 4

Family use English with the child:
Use English only (1) Use English and other languages (2)

Family use no English with the child:
Use one language (3) Use more than one language (4)

Please indicate languages used:

[5] POSITION OF CHILD IN THE FAMILY: **1 2 3 4 5 6** _____

[6] AGE OF MOTHER AT BIRTH OF FIRST CHILD**:**_____

[7] DOES THE CHILD OR PARENT HAVE ANY RECOGNISED SPECIAL NEEDS?

0 No Special needs
1 Child special needs
2 Parent special needs
3 Child and parent special needs

IF SPECIAL NEEDS, PLEASE SPECIFY: _____

SUMMARY

[8] PEDS general concern	Mark one: A B C D E
[9] PEDS language concern	Mark one: NO=0 YES=1
[10] WORD COUNT	Out of 100
[11] USE OF WORDS	Out of 10
[12] WORDS TOGETHER	Out of 2

WORDS CHILDREN SAY:
Children can understand words before they start to speak. We are interested in the words your child can SAY. This list does not have all the possible words children use, just some of those words. Please tick the words your child can SAY. If your child says the word differently (eg they say 'tar' instead of 'car') then tick the word anyway.

Baa baa	Hat	Sky	All gone
Meow	Necklace	Zoo	Cold
Woof woof	Shoe	Friend	Fast
Ouch/ow	Sock	Mummy/mum	Happy
Uh-oh/ oh dear	Chin	Person	Hot
Bear	Ear	Bye /bye bye	Last
Bird	Hand	Hi/hello	Tiny
Cat	Leg	Thank you	Wet
Dog	Pillow	No	After
Duck	Comb	Shopping	Day
Horse	Lamp/torch	Chase	Tonight
Aeroplane	Rubbish	Carry	Them
Boat	Tray	Pour	This
Car	Plate	Finish	Our
Ball	Towel	Fit	Us
Book	Bed	Hug/cuddle	Where
Game	Bedroom	Listen	Beside
Sandwich	Settee/sofa	Like	Down
Fish	Oven/cooker	Pretend	Under
Sauce	Stairs	Rip/tear	All
Cream cracker	Flag	Shake	Much
Meat	Rain	Taste	Could
Peas	Star	Think	Need to
Juice	Swing	Wish	Would
Milk	School	Gentle	If
COLUMN TOTAL:			

HOW CHILDREN USE WORDS

For each of the following questions, circle your answer: OFTEN, SOMETIMES or NOT AT ALL.

Does your child ever talk about past events? For example, a child who saw a funfair last week might later say 'funfair', 'clown', 'band'.

OFTEN	SOMETIMES	NOT AT ALL
2	1	0

Does your child ever talk about something that is going to happen in the future, for example, saying 'bus' or 'car' before you leave the house on a trip, or saying 'swing' when you are going to the park?

OFTEN	SOMETIMES	NOT AT ALL
2	1	0

Does your child talk about objects that are not present, such as asking about a missing toy, talking about an object s/he cannot see, or asking about someone who is not there?

OFTEN	SOMETIMES	NOT AT ALL
2	1	0

Does your child understand if you ask for something that is not in the room? For example, would s/he go to the bedroom to get a teddy bear when you say 'where's the bear?'

OFTEN	SOMETIMES	NOT AT ALL
2	1	0

Does your child ever pick up or point to an object and name the person the object belongs to, even when that person is not there? For example, a child might point to their brother Tom's shoe and say 'Tom', even when Tom is not there.

OFTEN	SOMETIMES	NOT AT ALL
2	1	0

PUTTING WORDS TOGETHER

Has your child started to put words together yet, such as 'more juice' or 'there doggie'?

OFTEN	SOMETIMES	NOT AT ALL
2	1	0

Appendix 2: SSLM format 2004

PROGRAMME NAME:_____NUMBER:_____

[1] CHILD IDENTIFIER _____

[2] BOY / GIRL (circle)
0 1

DATE OF SSLM–R _____CHILD DATE OF BIRTH_____

[3] CHILD AGE AT SSLM–R _____completed months

[4] FAMILY LANGUAGE BACKGROUND (Please mark one only)
1 Use English only
2 Use English and other languages
3 Use one language, not English
4 Use more than one language, no English

Please indicate languages used:

[5] POSITION OF CHILD IN THE FAMILY: **1 2 3 4 5 6 _____**

[6] EDUCATION LEVEL OF MOTHER OR CARER (please mark one only, highest achieved):
A None
B GCSE or equivalent [minimum school leaving age qualification]
C A level or AS level
D Further education including university
E Vocational skills
F Not known

[7] DOES THE CHILD OR PARENT HAVE ANY RECOGNISED SPECIAL NEEDS?
0 No Special needs
1 Child special needs
2 Parent special needs
3 Child and parent special needs
IF SPECIAL NEEDS, PLEASE SPECIFY: _____

SUMMARY

	Summary score sheet reference		*No concern*	*Concern*
A2 PLUS A3	**[8]**	**Language concern**	0	1 2 3 4
A4	**[9]**	**Parental concern**	0	1 2
B	**[10]**	**Word count**	Enter count out of 50 First words AND other words	
C	**[11]**	**Words together**	2 1	0

A. INTRODUCTORY QUESTIONS

1. *Do you have any questions about your child's talking?*

Comments for local programme use only

2. *Do you have any concerns or worries about how your child talks in words or sentences?*

Circle one: NO A LITTLE YES

0 1 2

COMMENTS:

3. *Do you have any concerns or worries about how your child understands what you say?*

Circle one: NO A LITTLE YES

0 1 2

COMMENTS:

4. *Do you have any concerns or worries about how your child is learning and growing up?*

Circle one: NO A LITTLE YES

0 1 2

Comments for local programme use only

B. WORD LIST

C. WORDS TOGETHER

Has your child started to put words together yet, such as 'more juice' or 'there doggie'?

NOT AT ALL SOMETIMES OFTEN

0 1 2

WORD LIST: WORDS CHILDREN SAY

Children can understand words before they start to speak. We are interested in the words your child can SAY. This list does not have all the possible words children use, just some of those words.
Please tick the words your child can SAY. If your child says the word differently (e.g. they say 'tar' instead of 'car') then tick the word anyway.

FIRST WORDS

1.	Mummy / mum		5.	Juice		9.	Cold	
2.	Bye / bye bye		6.	Ouch /ow		10.	Hug / cuddle	
3.	No		7.	Cat				
4.	Ball		8.	Thank you				**TOTAL**

OTHER WORDS

11.	Aeroplane		21.	Towel		31.	Fit		41.	Wet	
12.	Car		22.	Bed		32.	Like		42.	After	
13.	Book		23.	Settee / sofa		33.	Rip / tear		43.	Day	
14.	Milk		24.	School		34.	Shake		44.	This	
15.	Hat		25.	Friend		35.	Think		45.	Our	
16.	Shoe		26.	Person		36.	Gentle		46.	Where	
17.	Leg		27.	Hello / hi		37.	Fast		47.	All	
18.	Pillow		28.	Shopping		38.	Happy		48.	Much	
19.	Rubbish		29.	Carry		39.	Last		49.	Need to	
20.	Plate		30.	Finish		40.	Tiny		50.	If	
COLUMN TOTAL			**COLUMN TOTAL**			**COLUMN TOTAL**			**COLUMN TOTAL**		

Appendix 3: The comparison data set

Choice of cases

The comparison set of cases was drawn from each of the three implementations as follows:

A. A set of programmes from Rounds 1 – 3: These were early-established programmes contributing data to each of the three implementations. It could be argued that a focus on established programmes would maximise the chances of finding change over the three implementations, so these programmes were chosen as a core of the comparison subset. 70 out of 87 such programmes provided data for three consecutive years; and a further 5 programmes provided data for 2001 and 2003 or 2003 and 2004 only. The remaining 12 programmes could not be used because of lack of score forms (8 programmes) or lack of recoding time (4 programmes.)

B. 30 programmes from Rounds 4 – 5 were chosen randomly from the 55 Round 4-5 programmes which had provided data for both 2003 and 2004 and for which the child score forms were available. (Time did not allow recoding of all such 55 programmes.)

The sets A and B together provided 105 programmes for the comparison data set, as set out in Tables 46 and 47.

Table 46: Cases used for the comparison data set; by programme and round

Data source	Data used	Comment
70 programmes [R1-3]	3 sets of data	Forms missing (8), or lack of recoding time (4) for remaining 12 programmes.
5 programmes [R 1-3]	2 sets of data: 2001 and 2003 only, or 2003 and 2004 only	
30 programmes [R4-5]	2003 and 2004 only	Chosen randomly from the 55 programmes with available child forms.

Table 47: Cases used for the comparison data set in relation to full cohorts

	Number of cases counted out of 50	Total number of cases in original data set	Percentage of original data set used for comparison data set.
2001	924	1615	59%
2003	1302	2866	45 %
2004	1409	4694	30%
COMPARISON DATASET	3635 children		

Relation to full data sets

The relationship between the comparison data set and the full set of data over the three implementations is important for establishing the validity of the comparison set for making claims about the changes over time. For this reason, the comparison set was compared to the full data set, and also to the residual set of data collected over the three years but not used for comparisons.

Table 48 below sets out relevant key features of demographics: gender balance, the levels of English –only backgrounds, special needs levels, and sampling sources. Only the proportions of language backgrounds proved different for the comparison set, with more English – only backgrounds.

Table 48: Demographic features of the comparison data set in relation to those cases not used for comparative analysis

	2001, 2003,2004 Three years all cases	**2001, 2003,2004 Comparison file cases**	**2001, 2003,2004 Cases not used for comparison**
N	9177	3635	5542
GENDER			
% Boys (n)	**51.3 (n=4700)**	**51.4 (n=1862)**	**51.2 (n=2838)**
LANGUAGE % (n)			
ENGLISH	**77.5 (7105)**	**81.0 (2943)**	**75.1 (4162)**
BILINGUAL WITH ENGLISH	18.9 (1734)	15.5 (565)	21.1 (1169)
NON-ENGLISH	3.0 (272)	3.4 (125)	2.7 (147)
Unknown	0.6 (58)	0.1 (2)	0.1 (56)
SPECIAL NEEDS % (n)			
None	**94.7 (8667)**	**94.5 (3433)**	**94.4(5234)**
Child	3.4 (307)	4.0 (147)	2.9 (160)
Parent	1.8 (167)	1.4 (50)	2.1 (117)
Child + Parent	0.2 (14)	0.1 (4)	0.2 (10)
SAMPLES % (n)			
Representative	**37.3 (3423)**	**37.8 (1375)**	**37.0 (2148)**
Varied	12.4 (1141)	11.4 (415)	13.1 (726)
Sure Start only	**26.7 (2451)**	**25.0 (908)**	**27.8 (1543)**
Mixed	5.7 (521)	6.3 (228)	5.3 (293)
Unknown	17.9 (1641)	19.5 (709)	16.8 (932)

Word counts

Tables 49 and 50 explore the mean word counts by language background; first for the full set, then recounted cases only, and then the residual set of cases available but not recounted. The tables show how (unintentionally) the more able bilingual cases are used in the comparative data set. It has not been possible to test statistically the degree of differences between the comparative set and the residual set. (This is due to limitations of the data entry formats. Additional data coding time could add to this analysis.)

Table 49: Mean word counts of the comparison data set in relation to those cases not used for comparative analysis, by language background (2001 and 2003 data).

	2001 and 2003 all cases	**2001 and 2003 Comparison file cases**	**2001 and 2003 Cases not used for comparison file**
N	4481	2226	2255
Mean 100 word count	46.87	**47.48**	**46.27**
Mean 50 word count	na	25.53	Na
ENGLISH ONLY			
N	3611	1846	1765
%	80.5	**82.9**	**78.3**
Mean 100	48.235	**48.13**	**48.34**
Mean 50	na	25.72	Na
BILINGUAL WITH ENGLISH			
N	702	304	398
%	15.7	**13.7**	**17.6**
Mean 100	41.809	**44.28**	**39.92**
Mean 50	Na	24.42	Na
NON-ENGLISH SPEAKING			
N	163	74	89
%	3.7	**3.3**	**3.9**
Mean 100	38.687	**44.11**	**34.18**
Mean 50	Na	25.14	Na

In Table 49 both the Bilingual with English cases and the non-English speaking cases are under-represented by frequency but seem to contribute the more able cases, with the subsets of bilinguals having a higher word count out of 100.

Table 50: Mean word counts of the comparison data set in relation to those cases not used for comparative analysis, by language background (2004 data).

	2004 all cases	**2004 Comparison file cases**	**2004 Cases not used for comparison file**
N	4694	1409	3285
Mean 50 word count	25.53	**26.24**	**25.23**
ENGLISH ONLY			
N	3494	1097	2397
%	74.4	**77.9**	**73.0**
Mean 50	25.78	**26.26**	**25.56**
BILINGUAL WITH ENGLISH			
N	1032	261	771
%	22.0	**18.5**	**23.5**
Mean 50	24.8	**26.44**	**24.24**
NON-ENGLISH SPEAKING			
N	165	51	114
%	3.5	**3.6**	**3.5**
Mean 50	24.54	**24.78**	**24.43**

Again in Table 50 the Bilingual with English cases are under-represented by frequency and again seem to contribute the more able cases, with the subset of bilinguals having a higher word count out of 50. The sampling for non-English cases in the comparison file is evenly matched to the residual set.

Taken together over the three implementations, it is possible that this unintentional sampling of bilinguals influences the scores out of 50 for the comparison purposes.

Appendix 4: PQRS calculations

PQR and S are used as indicators of change. They are percentages calculated by apportioning the cohort of children into four categories. The two parameters involved are word count score (above and below a threshold score) and presence or absence of parental concerns about language. The four categories are made up of the four possible combinations of these two parameters.

The discussion here concerns the threshold score for word count scores. In the first and second implementations of the SSLM the threshold was set at 20 out of 100. This corresponded to approximately one standard deviation below the mean score. For the third implementation the shortened SSLM format was circulated to programmes, with a recommended threshold of 13 out of 50 for calculating the percentages PQR and S. The figure of 13 out of 50 was based upon a limited data set of 262 cases for which detailed analysis could be undertaken at that time.

The principal aim of using PQRS is to track changes from one year to another; this means comparability between years is important. One approach is to take only the first and second implementation cases, which have been recounted as if the 50 word list had been used (this is about 2225 cases out of 4481 first and second implementation cases). Then from 2001 onwards there is a set of data, which can be compared using any designated consistent threshold. This approach relies on the assumption that the recounted subset accurately reflects the full data set for that year. Unfortunately there is a slight bias in the recounted datasets, which may mean that the first and second implementations are not well represented by the recounted subset (see further Appendix 3.)

Another approach for comparing PQRS across years is to find a good match of threshold scores between the 100 and 50 word versions. Since the publication of the 50 word list in 2003, further analysis on 1841 cases from the first and second implementations has suggested that 12 out of 50 is the closest match to the level of 20 out of 100. The threshold of 12 allows data collected on a 50 word list to be compared to the PQRS percentages of the first and second implementations (calculated on a 100 word list, with a threshold of 20).

Threshold calculations

Data comes from 72 programmes which gave data for both of the first and second implementations, and where recounting was possible. 100 and 50 word count distributions are compared. (These are cases where scores are available both out of 100 and out of 50.)

First Table 51 shows 2001 data, of which 923 out of 1615 cases were analysed using the 50 word list. The distributions of the word count scores were compared. The aim was to match the levels of PQR and S for the analysed subset of 923 cases, when calculated both using 20 out of 100 and 12 out of 50.

Table 51: 2001 data: Comparisons of PQR and S using various thresholds.

	2001 BASELINE OUT OF 100 THRESHOLD 20	**Analysed subset Out of 100 Threshold 20**	**Analysed subset Out of 50 THRESHOLD 12**	Analysed subset Out of 50 THRESHOLD 13
DATASET	N= 1615 full set	**N=923**	**N= 923**	N=923
P	10.9	**11.2**	**10.8**	11.7
Q	10.7	**10.6**	**10.9**	10.1
R	8.6	**7.6**	**7.8**	9.3
S	69.8	**70.6**	**70.4**	68.9

The word count score of 20 or less out of 100 represented 18.7 % of the 2001 distribution. Matching this, the score of 12 or less out of 50 represents 18.6% of the distribution of the 2001 subset.

In a similar way for 2003 data, 917 out of 2866 cases were scored both out of 100 and out of 50, and their distributions compared. Again, PQR and S were matched for the thresholds 20 out of 100 and 12 out of 50 (see Table 52).

Table 52: 2003 data: Comparisons of PQR and S using various thresholds.

	2003 OUT OF 100 THRESHOLD 20	**Analysed subset Out of 100 Threshold 20**	**Analysed subset Out of 50 THRESHOLD 12**	Analysed subset Out of 50 THRESHOLD 13
DATASET	N= 2866 full set	**N=917**	**N= 917**	N=917
P	10.4	**9.6**	**9.4**	10.0
Q	10.0	**9.3**	**9.5**	8.8
R	8.3	**6.8**	**7.5**	8.9
S	71.3	**74.4**	**73.6**	72.2

The word count score of 20 or less out of 100 represented 16.3% of the 2003 distribution. Matching this, the score of 12 or less out of 50 represents 16.9% of the distribution of the 2003 subset.

So, for each of the first and second implementations, the threshold 12 out of 50 most closely matches the distribution centile of the threshold 20 out of 100. As a result the PQRS generated using the threshold 12 out of 50 for the analysed data most closely matches the PQRS generated using the threshold 20 out of 100 for the same subset.

In the case of the 2003 analysed subset, there is more variation between the subset and the full set for the PQRS figures than for 2001.

Copies of this publication can be obtained from:

DfES Publications, PO Box 5050, Sherwood Park, Annesley,
Nottingham, NG15 0DJ; Tel 0845 6022260; Fax 0845 6033360;
Textphone 0845 6055560; E-mail mailto:dfes@prolog.uk.com

Please quote reference: SSU/2005/FR/014