Learning Difficulties in the Primary School Years: Predictability from On-entry Baseline Assessment

Peter Tymms, Christine Merrell, Brian Henderson, Stephen Albone and Paul Jones

**Durham University,
UK.**

Email: p.b.tymms@dur.ac.uk

Abstract

Data from several thousand children on a teacher administered baseline assessment at the start of primary school in England in 1999 were linked to their test scores at the age of 11 (2006) when they were about to leave their primary settings and move to secondary school. The baseline assessment, which formed a single measure made up of sub-scales, was found to be a good predictor of attainment at the end of 7 years of schooling, correlating up to nearly 0.7 on a variety of outcomes. Unlike other studies, the best indicator of later achievement in mathematics, reading and science was early mathematics, and in particular the ability to identify numbers and do informal sums. Letter identification, which was found to be useful for prediction of attainment at age 7, was much less important at age 11. A wide range of additional variables, including, vocabulary, concepts about print and formal sums were found to improve the prediction, although it still remained difficult to reliably identify special needs at an early stage. Home background added a little to the prediction once controls had been made for the baseline measures, as did age and sex.

Multi level regression analysis was used to explore the variance associated with schools and pupils on attainment at age 11. Schools were found to be associated with about 15% of the residual variance once controls had been made.

The results from the analyses are discussed in relation to

a) The value of a range of variables as predictors of later achievement in the classroom context.
b) The identification of special needs and giftedness.
c) The impact of schooling.

Introduction

Data from the Performance Indicators in Primary Schools (PIPS) on-entry baseline assessment (BLA), collected in 1999 from pupils starting school in England at the age of 4 years, were linked to outcome measures of reading, mathematics, science, vocabulary and non-verbal ability at the end of primary school (age 11 years) in 2005 for more than one thousand cases. The paper explores the extent to which outcomes at age 11 years could be predicted from the baseline scores as well as home background measures. It also explores the identification of special needs and the apparent differences between schools.

There is a long tradition of baseline assessment which might be extended to measures of "school readiness" and even to intelligence. Within the English context, a flurry of work in the 1960s was aimed at identifying special needs at an early stage but the uncomfortable proportion of false positives led to a demise of interest. In 1998, there was a revival of interest as on-entry baseline assessment of children within the first few weeks of starting school became a statutory requirement (Wolfendale and Lindsay 1999). Schools selected a baseline assessment scheme to suit their own requirements from a list of ninety one accredited schemes. Later, in 2003, the Foundation Stage Profile was introduced and replaced the previous statutory arrangements. This was a single assessment system to be used throughout the Foundation Stage (for children aged 3 to 5 years), culminating in a final assessment of children at the end of the stage. In September 2008, the statutory assessment arrangements for young children were changed further and although the Foundation Stage Profile remained in place, the way in which the children were assessed comprised largely of practitioner observations of child initiated activities.

The PIPS BLA was designed as the starting point of a monitoring system for schools. Schools and in some instances Districts, paid an annual fee to participate in the system. They assessed their pupils and sent the data to the Centre for Evaluation and Monitoring (CEM) at Durham University where it was standardised and feedback on each pupil returned to the school. The aim of the PIPS BLA was to provide a reliable and valid baseline from which progress could be monitored and acted on by teachers. It originated in 1994 and, between 1999 and 2003, it was one of the accredited schemes. After the introduction of the Foundation Stage Profile, when these ceased to be a statutory requirement, many schools continued to assess their new pupils with the PIPS BLA and it is now used in around 4000 schools per year with large numbers in England, Scotland and Australia. It is also used on a more modest scale in Germany, New Zealand and the Netherlands; it has been translated into 8 languages. The assessment includes measures of name writing, vocabulary, ideas about reading, phonological awareness, letter identification, reading, ideas about mathematics, counting/numerosity, simple sums, more difficult maths problems including formal sums with symbols and number identification. The English version collects additional data which indicate whether English is the child's first language, the postcode of the child's home, sex and attendance at pre-school.

This paper concentrates on the prediction of later reading, mathematics, science and other cognitive outcomes from the baseline assessment administered on entry to school. Previous work (Blatchford et al 1987; Clay 1972; Riley 1994; Stuart 1995; Tymms 1999) has identified several useful predictors of later reading. Letter identification seems to be valuable and some have suggested that it is the best, single indicator. Phonological awareness, in various forms, has also been shown to be a reasonable predictor as have Concepts about Print and Vocabulary.

Although there has been much work in relation to the prediction of later reading only a little has been linked to mathematics (but see for example Bryant *et al*. 1990). Further, most of the work has been carried out with relatively short gaps between the initial and final tests, typically of less than three years.

Only one paper has been found which links a baseline assessment to outcomes at age 11 (Savage *et al*., 2007). The authors linked nine parts of a test for 5 year olds to measures of reading, maths and science at age 11. They reported that phonological awareness was a significant unique predictor of all outcomes after controlling for a baseline assessment, pupil background measures, early literacy skills and letter-knowledge. In the regression analyses with various outcomes about 30% of the variance was “explained” in total and the BLA (The Infant Index) “explained” around 7%. The extensive phonological awareness measure, whilst significant, accounted for just a few percent of the variance. The authors noted that the measure might be operating as a "more general measure of cognitive ability". This hypothesis is particularly convincing when the BLA accounted for so little of the variance.

Tymms (1999) explored the capacity of the PIPS BLA to predict reading and maths at age 7. Fourteen sub-tests were analysed and all were found to be significant predictors, and after correction for unreliability the great majority correlated with the outcomes at around the 0.7 level. The best predictors of reading were Letter identification, Number identification, Writing of the child’s name and Counting respectively. The best predictors of maths were Number identification, Letter identification, Counting and Writing of the child’s name respectively. Further factor analysis indicated the assessment largely measured a single construct although a second factor relating to the kinds of things that might be expected to be taught at school, such as identifying two digits numbers, doing formal sums and identifying the harder letters. A third factor which involved matching shapes was also separated out. As a result of that and further analyses, the PIPS BLA was refined.

Measures

The PIPS BLA is carried out within the first few weeks of pupils starting school and takes approximately 20 minutes per child. It comes in two forms: a computer-delivered version and a text version, where the teachers used a booklet with the child and record the results on paper. Only the data from the computer-delivered version are analysed in this paper.

To carry out the assessment the teacher sits with a pupil beside the computer and the teacher controls the mouse. The computer selects questions which are presented verbally through sound files and the teacher records the child's correct or incorrect response on-screen. If the child does not speak English then the instructions are translated into the child's language, provided a suitable adult who speaks the language is available. Even when translated, the target words for vocabulary and phonological awareness are the English words, as are the words in the reading sections. The computer program used in English schools has sound files for Bengali, Urdu and Cantonese.

Questions are presented in a series of sub-tests, and in each case except one, questions are asked in increasing difficulty until the child gets three wrong in a row or 4 wrong altogether. Algorithms within the computer program make the decisions automatically. The sub-tests are listed in Table 1.

The assessment is repeated at the end of the year. The program takes account of which questions a child answered correctly at the start of the year and adjusts the questions presented at the end of the year accordingly.

Table 1 Description of PIPS BLA Sub-tests

Writing	Writing – the child is asked to write his/her own name and the quality of writing is scored against examples.
Vocabulary	Vocabulary – the child is asked to identify objects embedded within a picture.
IAR	Ideas about reading – assesses many of the ideas found in Marie Clay's Concepts about Print.
Phonics	The items in these sections are measures of phonological awareness, which is abbreviated to phonics. Repeating Words – the child hears a word and is asked to repeat it. Rhyming Words – the child selects a word to rhyme with a target word from a choice of three options.
Letters	Letter identification – a fixed order of mixed upper and lower case letters.
Reading	Word recognition and reading. This starts with word recognition and moves on to simple sentences that the child is asked to read aloud. The words within these sentences are high frequency and common to most reading schemes. This is followed by a more difficult comprehension exercises which require the child to read a passage and at certain points select one word from a choice of three that best fits that position in the sentence.
IAM	Ideas about mathematics – assessment of understanding of the vocabulary associated with mathematical concepts.
Counting	Counting and Numerosity – the child is asked to count four objects. These disappear from the screen and then the child is asked how many objects they saw. This is repeated with seven objects.
Sums (informal)	Sums – addition and subtraction problems presented without symbols.
Sums (formal)	More difficult maths problems including sums presented with formal notation.
Numbers	Digit identification – single, two-digits and three-digits.

Table 2 shows the maximum number of items and reliability for each sub-test. The reliabilities for Writing, Letters, Numbers and Total are test-retest figures whereas the others are person reliabilities taken from Rasch analyses.

Table 2 Maximum scores and reliabilities of PIPS BLA sub-tests

	Max score	Reliability	Comment
Writing	5	0.61	
Vocabulary	23	0.84	
IAR	10	0.77	
Phonics	17	0.65	
Letters	27	0.92	U shaped distribution
Reading	99	0.50	Many did not reach this
IAM	7	0.81	
Counting	4	0.36	60% got all 4 correct
Sums (informal)	8	0.61	
Sums (formal)	13	0.40	Many did not reach this
Numbers	21	0.93	
Total	234	0.98	

The BLA was standardised on a nationally representative sample of English schools in 1999 with a mean of 50 and a standard deviation (SD) of 10. The sample that was available for this paper, which could be matched to later assessments, had a mean of 46.1 and a SD of 8.6 indicating a below average group by nearly half a SD and a slightly restricted range.

The outcome measures at the end of primary school (Year 6 in England) are mathematics, reading and science. Details of those tests are given in Table 3. Additional data, designed to assess developed ability, were collected on vocabulary and non-verbal ability (the Problems of Position (POP) test, Moseley, 1976). With the exception of the POP test, the assessments were developed by CEM for use by schools in the PIPS project. Information about the reliability of these outcome measures are reported as Cronbach's alpha: 0.93 for maths, 0.96 for reading, 0.89 for science, 0.90 for vocabulary and 0.94 for non-verbal (www.pipsproject.org, 2009). For the purposes of this paper the scores from the 5 assessments were reported separately and were also combined to give a Composite score which involved weighting the five outcomes according to the ratio of regression coefficients found when predicting an independent general outcome.

A larger number of schools carried out the vocabulary and non-verbal ability assessments than reading, mathematics and science. This was because some schools used the statutory tests of attainment rather than the PIPS tests. The numbers of pupils and schools whose scores from the BLA were matched to the outcome measures in Year 6 at age 11 are recorded in Table 3.

Table 3 Outcome Measures

	Max score	Reliability	N pupils	N schools
Maths	52	0.93	1091	82
Reading	104	0.96	1087	82
Science	47	0.89	1086	82
Vocabulary	36	0.90	6320	489
Non-verbal	35	0.94	6260	488
Composite	*	*	1065	82

Since the Composite was a weighted measure, there are no figures for the maximum score or reliability.

Analyses

The analyses were conducted in four parts. The first looks at the BLA measure itself. The second looks at predictions and correlations, the third at the identification of special needs and the fourth at school differences.

The BLA Measure

As noted earlier the assessment as a whole had high reliability, which is necessary for predicting later outcomes. The shorter sub-sections are of varying reliability but nevertheless provide useful detail about each pupil for teaching purposes. The assessment overall forms a single coherent scale well and this is shown by within the Rasch Measurement (Bond and Fox 2001) framework. A principal Components Analysis in Winsteps of the residuals indicates that a single measure explains 73% of the variance and 27% is left as unexplained variance. Just 0.8% of the variance was associated with the first contrast and it would seem to be associated with things that might be specifically taught at home such as the reading of passages and the doing of formal sums. A similar finding was reported in Tymms (1999) but the variance associated with the second contrast is very small. Essentially the BLA forms a goods single scale. This might lead to a supposition that it would not make sense to look at the component parts of the test. But the test is presumably fractal (Andrich 2007) in the same way that a measure of general knowledge might hold more specific information on knowledge of science which might hold information on knowledge of chemistry which might hold information on knowledge of organic chemistry and so on.

Further data were collected at the same time as the BLA and these included two important dichotomous variables, sex and English as an additional language (EAL). Attendance at pre-school sessions was dichotomised to show those who had been to nursery for three or more terms full time and others. Differences in terms of effect size are shown in Table 4. In general, the girls had slightly higher scores than boys on the BLA amounting to effect sizes of between 0.04 and 0.42 the highest being for name writing and the lowest for IAM and Counting. For those children for whom English was an additional language the average scores were lower by between a quarter and one and a two thirds of a standard deviation units. Not unexpectedly the largest difference was for vocabulary and the least for informal sums. The data indicate that those who had attended full time nursery for several terms had similar BLA scores than those who had not. Perhaps surprisingly the differences were small but it should be noted that the pupils who had attended nursery the most tended to come from more deprived backgrounds (r=0.2).

Table 4 Advantage of one group over another in Effect Sizes

	Sex (girls)	EAL	Nursery
Writing	0.42	-0.36	0.07
Vocabulary	0.06	-1.66	-0.10
IAR	0.10	-0.97	-0.02
Phonics	0.10	-0.55	-0.02
Letters	0.16	-0.40	0.05
Reading	0.16	-0.45	-0.02
IAM	0.04	-0.61	-0.00
Counting	0.04	-0.44	0.06
Sums (informal)	0.08	-0.88	-0.01
Sums (formal)	-0.01	-0.28	-0.05
Numbers	0.04	-0.34	0.07

Predictions and Correlations

This section is tackled in stages starting with the overall pattern and then progressively probing the relationships in greater detail.

Overall Pattern

Table 5 shows the correlations of the total score on the BLA with the Year 6 outcomes. As expected, the highest correlation, 0.63[1], was with the Composite score, but correlations of between 0.5 and 0.6 were found for the curriculum-based outcomes of mathematics, reading and science as well as for vocabulary, whereas the correlation with the non-verbal outcome was around 0.4.

In the same table, correlations with deprivation are given. These are derived from the pupils' postcodes linked to the 2001 census data (Norman 2001). The deprivation scale used was Carstairs (Carstairs and Morris1989). In some cases data on free school meal entitlement was available where postcodes were not recorded. In such instances an equivalent Carstairs figure was substituted for the missing values. Deprivation correlated at around the -0.2 to -0.3 level with all the outcome measures, although rather less with the non-verbal outcome.

[1] A correction for range restriction produces a figure of 0.69

Table 5 Correlations between the total BLA and Y6 outcomes

	Maths	Reading	Science	Vocabulary	Non-verbal	Composite
Total BLA	0.56	0.56	0.55	0.56	0.36	0.63
Deprivation	-0.20	-0.20	-0.27	-0.26	-0.12	-0.26

All figures are very significant statistically.
Deprivation data were available for 93 % of cases shown in Table 3.

Predictions and Correlations: Four Broad Areas of the Baseline

Having established some general relationships in the data the next stage was to explore the predictive power of the sub-scales of the BLA. Firstly, the four broad areas shown in the Table 6 were examined.

Table 6 Broad areas of the BLA

	Reliability	Included
Vocabulary	0.84	23 vocabulary items
Phonics	0.65	Repeats and rhymes
Early reading	0.75	Writing, IAR, letters and reading words sentences and passages
Early maths	0.78	IAM, counting, shapes, informal sums and number + formal maths

The correlations with the outcome measures are recorded in Table 7a. They varied between 0.21 and 0.62 and were generally highest for the Composite outcome and for the Maths component of the BLA. They were lowest for the Non-verbal outcome, never rising as high as 0.4 and were generally lowest for the Phonics part of the BLA again never getting as high as 0.4.

But the BLA measures varied in their reliabilities; the lower reliability of Phonics must contribute to the lower correlations but it is interesting to note the supremacy of maths over vocabulary and early reading where the reliabilities were similar.

Table 7a Correlations between areas of BLA and Y6 outcomes

	Y6 Outcomes					
BLA	Maths	Reading	Science	Vocabulary	Non-verbal	Composite
Vocabulary	0.32	0.35	0.39	0.50	0.22	**0.42**
Phonics	0.30	0.36	0.30	0.34	0.21	0.39
Reading	0.38	0.38	0.39	**0.43**	0.24	**0.44**
Maths	**0.59**	**0.53**	**0.53**	**0.49**	0.37	**0.62**
N	1082	1002	1001	5600	5560	1065

All figures are very significant statistically.
Correlations of 0.4 or higher in bold

It might be thought that including pupils whose first language was not English in Table 7a might distort the correlation. But Table 7b, is very similar to Table 7 and yet only includes those pupils whose first language was English.

Table 7b Correlations between areas of BLA and Y6 outcomes for children with English as mother tongue

	Y6 Outcomes					
BLA	Maths	Reading	Science	Vocabulary	Non-verbal	Composite
Vocabulary	0.36	0.37	**0.40**	0.46	0.23	**0.44**
Phonics	0.31	0.36	0.34	0.31	0.19	0.38
Reading	**0.46**	**0.45**	**0.45**	**0.43**	0.29	**0.51**
Maths	**0.61**	**0.55**	**0.55**	**0.47**	0.37	**0.63**
N	939	935	935	4954	4914	915

All figures are very significant statistically.
Correlations of 0.4 or higher in bold

The next step involved regression analysis in which the four BLA areas together with deprivation, sex and age at the time of the BLA were used to predict each outcome separately. Table 8a shows the results.

The interpretation of the coefficients must be done with care because of the colinearity of the cognitive predictors and their varying reliabilities. Nevertheless the following points can be made:

a) All outcomes were quite well predicted with multiple R values up to 0.68. The best was for the Composite outcome and the least good was for Non-verbal outcome (0.4)
b) Age was weakly, but statistically significantly, negatively related to the outcome. Presumably the higher baseline scores for older children starting school are not reflected in the outcome measures and this relationship shows in the negative regression coefficient.
c) Girls made a little less progress than boys in all areas.
d) The most important predictor for all outcomes except Vocabulary was Maths and for Vocabulary the predictors Vocabulary and Maths were almost equally strong.

e) Deprivation was significantly linked to most of the outcomes but the link was weak, the beta coefficient being around -0.1.
f) As noted earlier it might be thought that the children for whom English was an additional language might have distorted the underlying relationships in the data. But Table 8b produced very similar results.

Table 8a Regression, beta coefficients

BLA	Maths	Reading	Science	Vocabulary	Non-verbal	Composite
Vocabulary	.006	.065	.115	.264	-.018	.099
Phonics	.051	.136	.086	.038	.034	.099
Reading	.072	.061	.051	.124	.073	.091
Maths	**.519**	**.421**	.394	.240	.319	**.471**
Deprivation	-.074	-.066	-.138	-.079	-.016	-.104
Sex	-.173	-.021	-.131	-.074	-.151	-.137
Age	-.125	-.218	-.114	-.086	-.055	-.179
R	**0.62**	**0.59**	**0.60**	**0.58**	**0.40**	**0.68**
n	890	886	886	5013	4972	866

All coefficients significant (p<.01) unless font 8pt (p>.01)
Coefficients of 0.4 and above are in bold.

Table 8b Regression, beta coefficients: Children with English as mother tongue

	Maths	Reading	Science	Vocabulary	Non-verbal	Composite
Vocab	.062	.097	.137	.254	.030	.130
Phonics	.031	.132	.074	.033	.021	.084
Reading	.070	.055	.051	.123	.057	.095
Maths	**.527**	**.430**	**.409**	.253	.323	**.479**
Deprivation	-.068	-.062	-.121	-.084	-.022	-.097
Sex	-.157	-.003	-.128	-.085	-.158	-.122
Age	-.135	-.227	-.126	-.102	-.073	-.189
R	**0.64**	**0.61**	**0.61**	**0.56**	**0.41**	**0.69**
n	827	823	824	4381	4344	804

All coefficients significant (p<.01) unless font 8pt (p>.01)
Coefficients of 0.4 and above are in bold

Predictions and Correlations: Sub-scales

The next stage was to look at the correlations between each of the sub-scales of the BLA and the Year 6 outcomes. These are given in Table 9, which shows that all of the correlations were of modest magnitude. All were significant at least the .05 level, and those that were 0.4 or over are in bold. Surprisingly perhaps, none of the correlations with any of the outcomes for Reading or IAM got as high as 0.4. It should be noted that few children are actually reading at this stage, and that might be the reason for the lack of prediction. None of the correlations with the non-verbal outcome reached 0.4. The highest correlations tended to be those for number identification and informal sums and they got as high 0.56 and 0.53 for the Composite outcome respectively.

Table 9 Correlations between specific sub-tests of BLA and Y6 outcomes

BLA	Maths	Reading	Science	Vocab.	Non-verbal	Composite
Writing	0.32	0.30	0.30	0.30	0.27	0.35
Vocabulary	0.32	0.35	0.39	**0.50**	0.22	**0.42**
IAR	0.28	0.26	0.31	0.34	0.20	0.33
Phonics	0.30	0.36	0.34	0.33	0.20	0.39
Letters	0.36	0.39	0.34	0.39	0.23	**0.43**
Reading	0.30	0.29	0.30	0.27	0.15	0.35
IAM	0.30	0.30	0.32	0.32	0.19	0.34
Counting	0.37	0.30	0.30	0.25	0.21	0.36
Sums (informal)	**0.47**	**0.40**	**0.46**	**0.40**	0.29	**0.50**
Sums (formal)	0.21	0.13	0.20	0.17	0.14	0.21
Numbers	**0.51**	**0.47**	**0.42**	**0.40**	0.31	**0.52**

Correlations of 0.4 and above are in bold.
All correlation were very significant (p<.01)
n ≅1000 for 1st 3 and last columns n≅5000 for columns 4 & 5, but for Sums (formal) it was approx 330 for the first three columns)

Because the scales varied so much in reliability corrections were made to produce Table 10 (corrections were not made for the reliability of the outcome measures).

Table 10 Correlations between specific areas of BLA and Y6 outcomes, corrected for the reliability of the predictors

BLA	Maths	Reading	Science	Vocab.	Non-verbal	Composite
Writing	**0.41**	0.38	0.38	0.38	0.35	**0.45**
Vocabulary	0.35	0.38	**0.43**	**0.55**	0.24	**0.46**
IAR	0.32	0.30	0.35	0.39	0.23	0.38
Phonics	0.37	**0.45**	**0.42**	**0.41**	0.25	**0.48**
Letters	0.38	0.41	0.35	**0.41**	0.24	**0.45**
Reading	**0.42**	**0.41**	**0.42**	0.38	0.21	**0.49**
IAM	0.33	0.33	0.36	0.36	0.21	0.38
Counting	**0.62**	**0.50**	**0.50**	**0.42**	0.35	**0.60**
Sums (informal)	**0.60**	**0.51**	**0.59**	**0.51**	0.37	**0.64**
Sums (formal)	0.33	0.21	0.32	0.27	0.22	0.33
Numbers	**0.53**	**0.49**	**0.44**	**0.41**	0.32	**0.54**

Correlations of 0.4 and above are in bold

Table 10 shows that many more of the correlations are now above the 0.4 level. In fact, all but three of the items related to the Composite outcome were above 0.4. It is also interesting to note that even with the corrections; letter identification was consistently behind number identification except when predicting vocabulary when they were equal. The best predictive measures were still mathematical (numbers, counting and sums (informal)) for all outcomes except vocabulary which was best predicted by vocabulary.

The next analyses involved regressions in which the sub-tests were used as predictors alongside deprivation, sex and age. The English as an Additional Language variable was excluded because it had earlier been shown to make little difference to the regression coefficients and its inclusion would involve some loss of data. The Sums (formal) measure was omitted because it has been shown to be a relatively poor predictor and because its inclusion would have cut the number of available cases to a third.

Table 11 indicates that numbers was the most important predictor of the outcomes except for vocabulary and non-verbal measures. The ability to do sums on entry to school was the only other of the sub-tests to be significant at the 1% level for all outcomes. Writing was only significant in predicting the non-verbal outcome. Vocabulary was significant in four out of six regressions and Phonics was significant in predicting reading, vocabulary and the composite outcomes. IAR and Reading were never significant whereas IAM was significant for three outcomes as was Counting.

Deprivation, sex and age produced similar patterns to those seen earlier in Table 8.

Table 11 Regression analyses: Beta coefficients

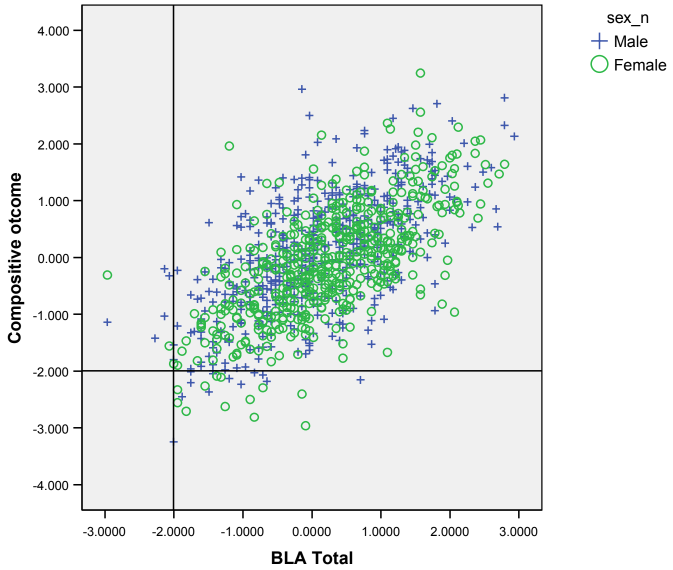
	Maths	Reading	Science	Vocab.	Non-verbal	Composite
Writing	.060	.009	.000	.004	.143	.033
Vocabulary	.043	.109	.122	.264	-.008	.138
IAR	-.039	-.057	.002	-.006	-.007	-.044
Phonics	.045	.145	.069	.050	.043	.101
Letters	.003	.067	.008	.127	.006	.059
Reading	.080	.020	.040	.033	.002	.061
IAM	.071	.081	.115	.066	.018	.079
Counting	.116	.070	.047	.009	.065	.088
Sums (informal)	.168	.112	.184	.097	.108	.168
Numbers	.297	.278	.216	.123	.123	.263
Deprivation	-.103	-.086	-.159	-.095	-.034	-.132
Sex (female)	-.174	-.011	-.117	-.072	-.171	-.140
Age	-.129	-.223	-.114	-.080	-.072	-.182
R	**0.62**	**0.59**	**0.60**	**0.56**	**0.37**	**0.68**
n	784	781	781	4211	4179	2624

All coefficients significant (p<.01) unless font 8pt (p>.01)
Coefficients of 0.4 and above are in bold
Formal sums were not included because it would have reduced the sample by more than 50%

Identifying special needs

It might naively be thought that the BLA could be used by teachers to identify children with special needs. Often a cut-off of two standard deviations below the mean of an assessment is used and for the BLA this is shown in Diagram 1 which also indicates the breakdown between males and females. The horizontal line shows a similar cut-off on the Composite outcome. The bottom left hand quadrant created by the lines shows the true positives. The top left and the bottom right quadrants respectively indicate false positives and false negatives. There were no true positives.

Diagram 1: Scattergram of Composite outcome against BLA



The situation is underlined with the following data. There were 30 cases occupying the bottom 5% on the BLA. Of these about a third (13) were still in the bottom 5% in Y6. About a fifth (6) were in the next 5% and a just 2 in the next 5%.

At the other end of the distribution similar figures were found. The predictive power of the BLA was as good for the less able as it was for the gifted.

A BLA carried out on entry to school will have been influenced by all sorts of factors such as the effort at home to teach letters, the degree to which the child has settled at school and the amount and quality of stimulation. None of these were measured. But by re-assessing at the end of the first year at school the influence of some of these factors can be reduced and a better prediction can surely be made. If nothing else, a second assessment will reduce the uncertainties of measurement. Predicting the Composite outcome from the total scores at the start and end of the year produces a multiple R of 0.71. This is better than the 0.63 based on the start of year alone. It still does not allow for anything like a firm identification of special needs or giftedness. However, the BLA could start the process of identifying special needs. It alerts teachers to children at risk at an early stage of their schooling, and their progress can be carefully monitored.

School Differences

This section tackles three questions:

a) Are the conclusions reached this far threatened because the clustering of pupils in schools was ignored?
b) To what extent can primary school differences at age 11 be understood in terms of the pupils who joined at age 4?
c) Do the correlations between the BLA and attainment at 11 vary across schools?

a) The effect of clustering

To deal with the clustering of pupils within schools, it is appropriate to run multilevel models. The multilevel models which follow start with null models in which the variance of the outcome is split between pupils and schools. These are shown in Table 12. Around a quarter to one third of the variance was found to be associated with the schools.

Table 12 Null multi-level models

	Maths	Reading	Science	Vocab.	Non-verbal	Composite
Fixed						
Cons	46.6 (0.55)	46.3 (0.59)	47.0 (0.60)	48.1 (0.23)	48.5 (0.22)	48.8 (0.65)
Random						
Pupil	65.9 (2.9)	63.0 (2.8)	62.9 (2.8)	66.7 (1.2)	74.5 (1.4)	65.9 (2.9)
Schools	14.4 (3.5)	18.1 (4.1)	18.8 (4.3)	16.4 (1.5)	13.5 (1.4)	23.2 (5.1)
% at sch.	18%	22 %	23%	20%	15%	26%

The numbers in parentheses are standard errors on the coefficients

In Table 13 the models are extended so that the outcomes are predicted using the total BLA score as well as deprivation, sex and age.

Table 13 Multi-level models with fixed predictors

	Maths	Reading	Science	Vocab.	Non-verbal	Composite
Fixed						
Cons	31.9 (3.5)	39.6 (3.5)	33.8 (3.6)	32.3 (1.5)	37.7 (1.7)	35.5 (3.4)
BLA	0.64 (0.03)	0.63 (0.03)	0.59 (0.03)	0.60 (0.01)	0.43 (0.01)	0.74 (0.03)
Depriv.	-0.06 (0.09)	0.05 (0.09)	-0.20 (0.10)	-0.16 (0.04)	-0.01 (0.04)	-0.02 (0.09)
Sex	-3.27 (0.44)	-0.40 (0.43)	-2.50 (0.44)	-1.35 (0.19)	-2.90 (0.22)	-2.64 (0.42)
Age	-3.18 (0.77)	-5.1 (0.76)	-2.98 (0.76)	-2.46 (0.33)	-1.68 (0.38)	-4.62 (0.73)
Random						
Pupil	45.4 (2.1)	43.2 (2.0)	45.9 (2.3)	48.5 (0.9)	63.6 (1.2)	38.9 (2.6)
School	7.0 (1.9)	9.5 (2.4)	7.9 (2.1)	7.4 (0.8)	10.9 (1.2)	10.8 (2.6)
% At school	13%	18%	15%	14%	15%	22%
Explained at pupil level	31%	31%	27%	27%	15%	41%
Explained at school level	51%	48%	58%	55%	19%	53%

Because the BLA and outcome measures had the same SD (10) the coefficient in the multilevel models can be interpreted as correlations and they are fairly close to those recorded in Table 4. Although close, they are a little higher presumably because clustering is now taken into account and because the effects of deprivation, sex and age are taken into account.

b) "Explaining" school differences

Around a third of the variance at the pupil level is "explained" by the control variables. This figure varied from 15% for non-verbal ability to 41% for the composite outcome. About half of the school level variance is "explained". In other words the starting points of pupils go a considerable to help understanding the differences between the average attainment levels at the end of primary school education. Now around 15% of the variance is associated with the school. This varies according to outcome measure although since the error on the school variances is around 10% it would be hard to claim that there is a great difference between outcome measures.

C) Do the associations between the BLA and outcome measures vary from school to school?

The models in Table 13 were extended to allow the BLA slope to vary but the slope variance was not significant except for vocabulary and the composite outcome, which included vocabulary. This provides evidence that the correlation between the BLA and vocabulary was different in different schools. There was not a big difference but it did vary.

Summary

The analyses presented above indicate that the PIPS BLA with children who are just four and a half years old is associated with around 50% of the variance in the outcome measures of pupils leaving primary school at age 11 years. More detailed analysis indicated that the best predictions for reading, mathematics and science came from the mathematics area of the BLA and in particular from sums (informal) and number identification. Letter identification was less good but still significant. Additionally, vocabulary, phonics and reading did correlate with the outcome measures at around the 0.3 to 0.4 level. To some extent, the assessment was compromised by the necessary time limitations imposed on it but if the reliabilities of each of the sub-units are taken into account the same conclusions remain.

Regression analyses which include deprivation, sex and age, showed that numbers and sums (informal) were significant predictors of all outcomes. Letter identification did not add to the prediction of reading, maths or science. But phonics did add weakly and significantly to the prediction of reading and vocabulary.

Even a correlation of 0.63 between a BLA and its outcome is not sufficiently good to identify children with special needs with any degree of surety. The proportion of false positives is around two-thirds. The same figures were found when trying to identify gifted children. Re-assessing children at the end of the first year at school improves the prediction with a correlation of 0.71, but this improvement does not circumvent the difficulty of identifying children with special needs at a very early stage. This, of course, does not undermine the use of a BLA as a starting point for identifying special needs, nor its value in starting the process of tracking child's progress.

An analysis using multilevel models indicated that the BLA in conjunction with deprivation, age and sex accounted for more than half of the school level variance and up to 40% of the pupil level variance. After controls, between 13-22% of the variance was associated with the school; a figure which is commonly found in school effectives studies for secondary school although, we are no aware of any study which has reported a figure for the start to end of primary schooling hitherto.

The multilevel models also gave some indication that the links (correlations) between the BLA and outcomes at age 11 are stable across schools although they did vary significantly when the outcome was vocabulary and a composite measure, which included vocabulary

Conclusion

This study adds to evidence for of the validity of the PIPS BLA; it is possible to use an objective approach to produce good measures of four year old children. It also raises some interesting parallels and contrasts with other studies.

Previous work has indicated that letter knowledge is one of the best predictors of reading up to the age of age 7 but this paper indicates that this no longer held once the children reached the age of 11 when letter knowledge appears to be a rather weak predictor. Number identification was the stronger predictor and this held for mathematics and science outcomes as well. The ability to do informal sums on entry to school was the second best predictor across the same three outcomes. Perhaps mathematics remains a purer measure of development than early reading skills because of parents to encourage reading. Another possible reason is that informal sums and number identification develop more naturally as a child matures. Reading is a human construct and only starts to show up with a certain level of development and exposure. Whatever the reasons it is clear mathematics was a better predictor of not just maths but of reading and science.

As expected there were sex differences, although they were fairly small, amounting to around 0.2 of a standard deviation in favour of the progress made by boys. It might be odd to note that the females were doing less well but the negative coefficients in the regression analyses appeared after controls for the BLA where they were doing rather better than males.

Age was significantly related to outcomes and the negative coefficients were presumably due to the older children doing particularly well at the start of school. Age is not so important at 11 when one year is 10% of child's life as it is at the start of school when one year is 25% of the child's age. Further, although the amount of schooling is clearly related to attainment, that was the same for all in the study. By the time children were at the end of primary school they had engaged in 7 years of education.

The finding from the multilevel analysis that the BLA accounted for a substantial proportion of school differences comes as no surprise. But the similar proportion of school level variance associated with the school for vocabulary and non-verbal ability at age 11 when compared with reading mathematics and science is a little surprising. It might have been thought that schools would have most impact on subjects which are specifically taught. But there was also an indication that the link between the BLA and vocabulary at age 11 varied across schools. There is evidence that schooling in general increases IQ (see for example Winship and Korenman 1997). Could it be that schools are not just varying in their capacity to influence the attainment levels of their children but also the developed abilities?

References

Andrich, D. (2007). On the fractal dimension of a social measurement

Bond, T.G. and Fox, C.M. (2001). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Mahwah, NJ, Lawrence Erlbaum Associates.

Blatchford, P., Burke, J., Farquhar, C., Plewis, I. and Tizard, B. (1987). Associations Between Pre-school Reading Related Skills and later Reading Achievement. *British Educational Research Journal.* **13**(1): 15-23.

Bryant, P.E., MacLean, M., Bradley, L.L. and Crossland, J. (1990). Rhyme and Alliteration, Phoneme Detection, and Learning to Read. *Developmental Psychology* **26**(3): 429-438.

Carstairs, V. and R. Morris, R. (1989). Deprivation, mortality and resource allocation. *Community Medicine* 11: 364 - 372.

Clay, M. M. (1972). *The Early Detection of Reading Difficulties: A Diagnostic Survey*. Auckland, Heineman.

Moseley, D. (1976). *Helping with learning difficulties*. Milton Keynes, Open University.

Norman, Paul (2001) Cathie Marsh Centre for Census and Survey Research (CCSR), 2001 Deprivation Indexes: Townsend and Carstairs scores [computer files]. ESRC/JISC Census Programme, Census Dissemination Unit, MIMAS (University of Manchester) http://130.88.204.31/cdu/2001/help/faq.htm#deprivation

Savage, R., Carless, S. and Ferraro, V. (2007) Predicting curriculum and test performance at age 11 years from pupil background, baseline skills and phonological awareness at age 5 years *Journal of Child Psychology and Psychiatry* **47** (7) 629-739

Stuart, M. (1995). Prediction and Qualitative Assessment of Five and Six Year Old Children's Reading: A Longitudinal Study. *British Journal of Educational Psychology*(65): 287-296.

Tymms, P. (1999). *Baseline Assessment and Monitoring in Primary Schools: Achievements, Attitudes and Value-added Indicators*. London, David Fulton Publishers.

Tymms, P.B. (1999). Baseline assessment, value-added and the prediction of reading. *Journal of Research in Reading* **22**(1): 27-36.

Tymms, P., Merrell, C. and Henderson, B. (1997). The First Year at School: A quantitative investigation of the attainment and progress of pupils. *Educational Research and Evaluation* **3**(2): 101-118.

Tymms, P., Merrell, C. and Henderson, B. (1997). *The progress of children during the first three years at school.* European Conference of Educational Research, Frankfurt, Germany.

Winship, C. and Korenman, S. (1997). Does Staying in School Make You Smarter? *Intelligence, Genes and Success: Scientists Respond to The Bell Curve*. B. Devlin, S. E. Fienberg, D. P. Resnick and K. Roeder. New York, Copernicus.

www.pipsproject.org (2009) website accessed on 3rd August 2009.