Department for Education

Performance Indicators in Primary Schools: A comparison of performance on entry to school and the progress made in the first year in England and four other jurisdictions

Research report

June 2014

Peter Tymms, Christine Merrell, David Hawker & Frazer Nicholson - CEM Centre, Durham University

Contents

List of figures

List of tables

Executive summary

This report describes secondary analysis of existing data from the Performance Indicators in Primary Schools (PIPS) On-entry Baseline and Follow-up assessments, as used in five jurisdictions: England, Scotland, New Zealand and one State and one Territory in Australia. Children across each jurisdiction were assessed using the Baseline assessment on entry to school and using the Follow-up assessment at the end of their first year of formal education.

Firstly, the report establishes the validity of the PIPS measures of cognitive and non-cognitive development for particular groups of children, and to produce a national average of what children can do on entry to school in each country.

For the first time, using a nationally representative sample and objective[1] data, what children know and can do when they start school in England is described. The report then sets out the links between these measures and key background variables including gender, age, deprivation, school attended, pre-school experience and jurisdiction.

Finally, it draws conclusions regarding the feasibility of a robust international comparative study of children starting school and progress during their first year.

The cognitive measures of *early reading*, *early mathematics*, *vocabulary* and *phonological awareness* were found to work well in England for the full target age range, the socioeconomic spectrum and both sexes. Comparisons were more challenging for children whose first language was not English or who may have been encountering English for the first time when they started school. The measures of *early reading* and *early mathematics* were found to work consistently across the five jurisdictions (after some items were dropped), and therefore it is possible to make direct comparisons.

Children in all five jurisdictions were found, on average, to make substantial progress by the end of their first year at school in all the cognitive domains. They also made considerable progress on the non-cognitive measures of personal, social and emotional development.

The baseline assessments revealed that, of the five jurisdictions, children starting school in England had the lowest overall scores when they started school. However, age for age the English children were in line with children from Scotland and New Zealand for *early mathematics* but significantly behind their peers from Australia. For *early reading*, children from England had the lowest scores at the start of school, age for age and this was statistically significant. Scotland was slightly ahead of England and children from the Antipodes ahead of both England and Scotland.

The follow-up assessments showed that progress in *early reading* and *early mathematics* during the first year at school varied widely from pupil to pupil and from school to school. The

---

[1] In the sense that the results are not dependent on who collected the data

study confirmed previous research (including studies using PIPS, for example Tymms, Jones, Albone and Henderson, 2009) which found that children's progress is much greater during one year at this age than at any other age during their school career. It also found that their progress is more dependent on the school attended than has traditionally been found in most other studies of schooling internationally. For *early mathematics*, the progress was generally similar across jurisdictions whereas for *early reading*, the progress was greatest in England.

An international comparative study of children starting school faces two challenges. The first is that children start school at different ages and the second is that languages and context vary from country to country. This report has shown that it is possible to make comparisons of what children can do at an early age across English-speaking countries, using the PIPs assessments, despite the differing ages.

Introduction

This report looks at the cognitive and non-cognitive development of children starting school and their progress to the end of the academic year. The instruments used for the assessment were the Performance Indicators in Primary Schools (PIPS) On-Entry Baseline and Follow-up assessments and the data came from England, Scotland, New Zealand and one State and one Territory in Australia. The first data were collected at the start of the academic year in England in 2011 and again at the end of the academic year in 2012. The Australian and New Zealand data were collected during 2012 and the Scottish data were collected during the year 2012/13.

The analysis required clean and nationally representative datasets from the different jurisdictions, therefore the first stage of the analysis was to establish these from the data available in each jurisdiction. Appendix 1 describes the procedures used to establish the datasets.

Once a representative sample was established for England, the data were used to explore what children know and can do when they start school in England and create the performance scales which are used in later analyses.

In order to help assess the quality of the data and to provide a measure which places both items and children on a comparable scale in order to be able to make judgements and comparisons, a particular approach known as Rasch measurement was used.

The extent to which the assessment instruments behave differently in different contexts and where it was possible to equate scales was first investigated. The relative progress made by children from the five jurisdictions was compared. Progress is firstly presented in graphical format and then presented using multi-level models with children nested within schools. Jurisdictions are identified by dummy variables in the models. The impact of other variables is also considered.

The report finally considers whether it is possible to run a robust international project designed to study children’s development at the start of school across different countries. The judgement takes account of the analyses within this report as well as practical and technical issues. It is concluded that such an international study is quite feasible.

Background to the Performance Indicators in Primary Schools (PIPS) monitoring system

The Performance Indicators in Primary Schools (PIPS) monitoring system began nearly 20 years ago in England with the intention of providing baseline data for primary schools in order to be able to generate confidential, nationally based, value-added scores at a later stage in the children's education. It quickly shifted to providing information to teachers about what their pupils knew and could do, how their performance compared against norms and their progress throughout primary school. The starting point for the system was the PIPS On-entry Baseline Assessment. The assessment was initially used by a small number of schools in the north-east of England and Solihull but spread rapidly by word of mouth and with backing from the National Association of Head Teachers. As its use by teachers increased, they asked for a repeat assessment at the end of the Reception year in order to be able to record the substantial progress made by children during their first year at school, and it was thereafter referred to as the PIPS On-entry Baseline and Follow-up.

The use of the PIPS On-entry Baseline and Follow-up (PIPS) has gradually expanded and the assessments have been translated and adapted into a dozen languages. It has now been used to assess more than a million children across several countries.

Full documentation outlining the PIPs assessments are available on the Centre for Evaluation and Monitoring's website[2]. In short, the assessment has two major parts to it: the first assesses cognitive development; and the second, non-cognitive development. The first part is designed to assess those aspects which are the best predictors of later success or difficulty (Tymms, 1999a and 1999b; Tymms, Brien, Merrell et al, 2003; Tymms, Merrell, Henderson et al, 2012). It is a computer-based assessment and the content is very comprehensive but no individual child sees all of it, because it is presented in a series of sequences with stopping rules. This adaptive nature means that a very wide range of children can be assessed in an efficient, reliable and enjoyable way. The cognitive section is made up from sub-units which assess *Early Mathematics*, *Early Reading*, *Picture Vocabulary* and *Phonological Awareness*. There is also a *Short Term Memory* Assessment but that is optional and the results are not reported here.

The non-cognitive section assesses personal social and emotional development and is based on teachers' observations of their pupils' behaviour in the school setting. There is also an additional section for the identification of inattention, hyperactivity and impulsivity, which is based on the Diagnostic and Statistical Manual of Mental Disorders (Version IV) published by the American Psychiatric Association (1994).

[2] http://www.cem.org/pips-baseline/introduction

This report focusses predominantly on the cognitive element of the assessment and reports some of the findings from the personal, social and emotional development section. It starts by concentrating on England and then broadens out to consider other jurisdictions.

Establishing representative samples of PIPS data for England, Scotland, New Zealand and two Territories/States in Australia

The data which were available from England, Scotland, New Zealand and the two Australian territories/states (Australia A and Australia B) were not collected as representative samples. Rather, representative samples for this project were defined. The way that this was done is described in Appendix 1.

Data available from the samples

The major part of this report is the analysis of the assessment of children's cognitive development. The numbers of pupils in the representative samples at the start of the year are shown below inTable 1. Table 2 gives the numbers at the end of the academic year.

Table 1: Numbers of pupils for whom start of year cognitive development data were available

Jurisdiction	Early Maths	Early Reading	Picture Vocabulary	Phonological Awareness
England	6,983	6,983	6,983	6,983
Scotland	6,623	6,623	6,623	6,623
New Zealand	847	847	847	847
Australia A	3,457	3,457	3,421	3,457
Australia B	1,313	1,313	1,311	1,313

Table 2: Numbers of pupils for whom end of year cognitive development data were available

Jurisdiction	Early Maths	Early Reading	Picture Vocabulary	Phonological Awareness
England	5,939	5,939	5,939	5,939
Scotland	6,627	6,627	6,627	6,627
New Zealand	594	594	594	594
Australia A	3,205	3,205	3,205	3,205
Australia B	1,228	1,228	1,228	1,228

Schools could, optionally, complete the non-cognitive part of the assessment, which contains items designed to measure personal, social and emotional development (PSED) and the behaviour section which assesses characteristics associated with inattention, hyperactivity and impulsivity (behaviour). The number of pupils for whom data were available is given below.

Table 3: Pupils for whom start of year non-cognitive development and behaviour data were available

Jurisdiction	PSD	Behaviour
England	1,566	0
Scotland	669	0
New Zealand	65	0
Australia A	0	103
Australia B	0	75

Table 4: Pupils for whom end of year non-cognitive development and behaviour data were available

Jurisdiction	PSD	Behaviour
England	1,264	925
Scotland	726	641
New Zealand	17	12
Australia A	0	172
Australia B	0	52

Data were available on various background variables. These are listed below as percentages of the representative sample.

Table 5: Additional background data available

Jurisdiction	School	Class	DoB	Free meal entitlement	Terms at Nursery	Terms at Playgroup	Flagged for English as an additional language	Ethnic origin	Flagged for Special educational needs (communication, deafness, cognitive, behaviour or physical)	Deprivation (IDACI for England, SMID for Scotland)
England	100	100	90	15	34	7	85	13	7	100
Scotland	100	100	100	40	0	2	22	2	4	100
New Zealand	100	100	89	0	0	0	55	66	5	0
Australia A	100	100	100	0	0	0	100	85	4	0
Australia B	100	100	100	0	0	0	100	23	8	0

The PIPS assessment

The PIPS assessment has more than 200 items within the cognitive part. Some of those are individual items located within long reading passages whilst most are standalone items. They are all organised into a series of sub-assessments corresponding to constructs relevant to early reading and early mathematics and each of the items has been shown to be predictive of later success in literacy and numeracy (Tymms 1999b, Tymms et al 2012, (Table 6).

Table 6: Cognitive constructs used in PIPS

Construct	Description
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.
Ideas About Reading	Ideas about reading – assesses many of the ideas found in Marie Clay's Concepts about Print. (Clay 1972)
Phonological awareness	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.
Early 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 exercise which requires 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.
Ideas About Mathematics	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.
Digits	Digit identification – single, two-digits and three-digits.
Shapes	Identification of a variety of geometric shapes
Number	Number manipulation – the child is asked how many more or less a number is than a target
Sums A	(Informal) Sums – addition and subtraction problems presented without symbols.
Sums B	(Formal) More difficult maths problems including sums presented with formal notation.

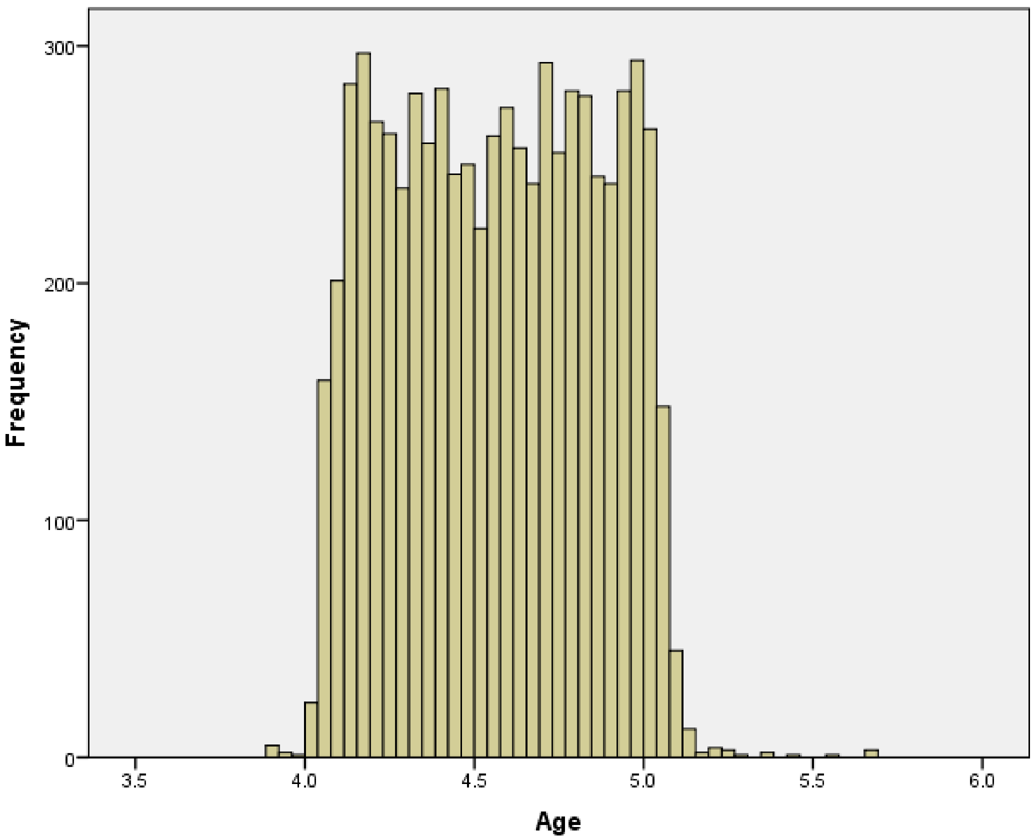
Pupils are shown each section of the assessment on-screen. They start with easy items and continue on to progressively harder ones within that section until they get three wrong in a row or four wrong altogether. They then move to the next section and so on.

The non-cognitive part of the assessment is completed by teachers based on their knowledge of working with the children. There are two parts. One concerns personal, social and emotional development and this is made up of 11 items, each of which has a rating scale. The items and scales are set out in Appendix 6. The other relates to behaviour and, since the data are collected at the end of the year in England, it is not reported here, but several academic papers have been published which are based on this information (Merrell, Styles, Jones et al, 2013; Sayal, Owen, White et al, 2010; Merrell and Tymms, 2001).

Background variables for England

The first part of this report explores the PIPS data from England, how children vary when they start school and their progress during the first year. In order to contextualise the information and to make fair comparisons, four key background variables are used: Age, gender, English as an Additional Language (EAL) and deprivation (home postcode used to link to Income Deprivation Affecting Children Index - IDACI). Further information is given below.

Figure 1: Distribution of children's ages at the start of school



Age

The vast majority of pupils started school in September (99.5%) with a few starting in January and even fewer after Easter. The distribution approximates a rectangle with a very small number joining after the age of five and a similar number aged just under four[3].

Gender

The sample was made up from 48.5% girls and 51.5% boys.

English as an Additional Language (EAL)

1,110 pupils (15.9%) were identified as having English as an Additional language[4].

Deprivation

The distribution of deprivation measures (IDACI scores) is shown in Figure 2. This is based on home postcodes and a high score indicates high deprivation. The distribution is very similar to the distribution for all children staring school in England[5].

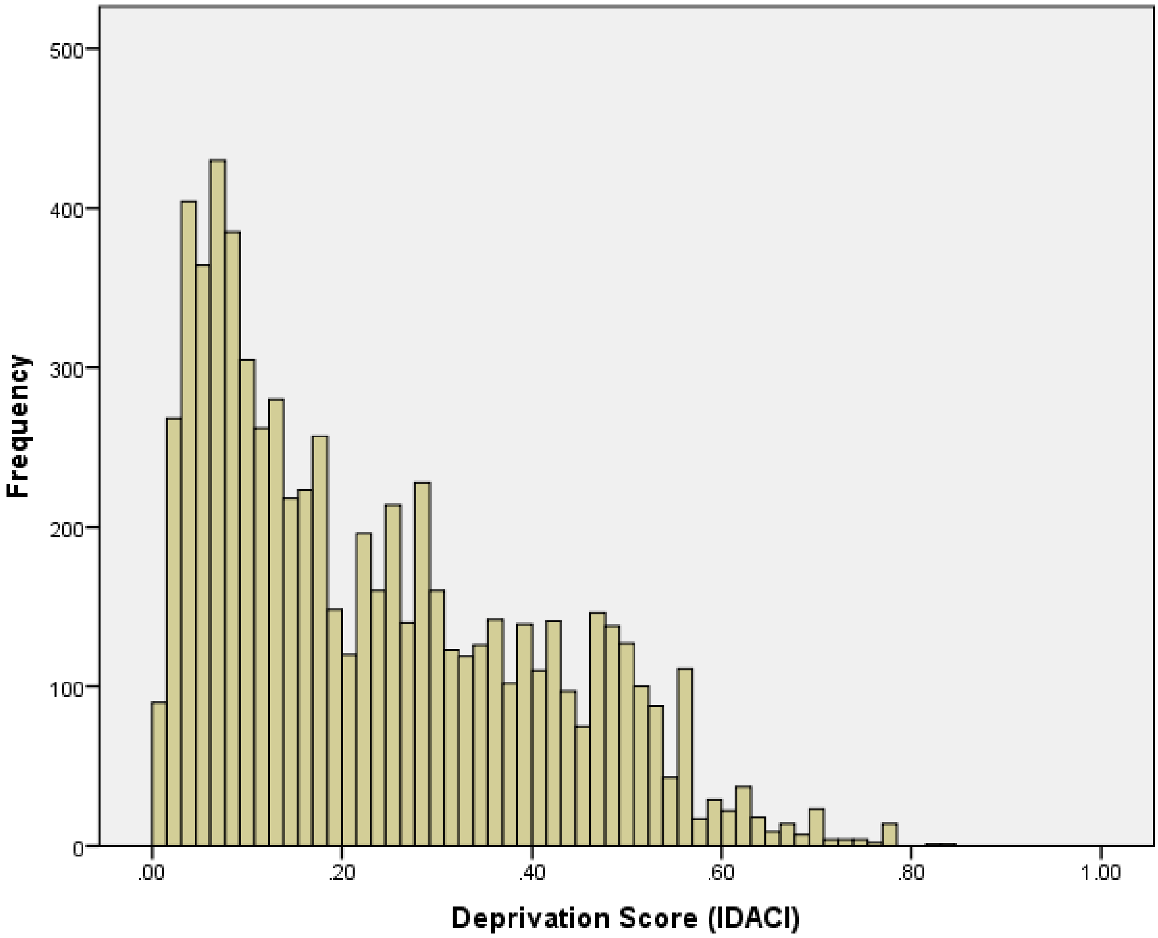
---

[3] Data were not available at the time of writing to compare the sample to the full population of children starting school

[4] Data were not available at the time of writing to compare the sample to the full population of children starting school

[5] The PIPS sample is based on 6,986 cases and has a mean of 0.23 and a Standard Deviation of 0.17. The national data are based on 610, 023 cases. Their mean is 0.24 and their Standard Deviation is 0.18

Figure 2: The distribution of deprivation measures (IDACI) based on home postcodes. High score indicates high deprivation



Establishing the measurements

PIPS is now nearly 20 years old and it has given rise to almost 50 academic articles. These papers and the associated research have gone a long way to establish the reliability and validity of the assessment, and some key figures are given in Appendix 2. One important point to note is that in order to establish the consistency of PIPS, a group of children, chosen at random, were reassessed by a researcher after they had been assessed by their teachers. The two results were almost identical[6]; the assessment is extremely reliable. A second key point is that children have now been followed up for many years and links between PIPS and later assessments have been published (Tymms, Jones, Albone et al, 2009; Tymms, Merrell, Henderson et al, 2012). Figures are given in the appendix which show good predictive validity[7].

---

[6] The correlation between the total scores, excluding phonological awareness, was 0.98 for 29 pupils

[7] The correlation to the end of Key Stage 1 average level was 0.66

The data were analysed using a technique known as Rasch measurement, which very usefully puts items and individuals on the same scale. This presents a clear visual illustration of what children know and can do. The approach is also attractive because it takes account of missing data which occurs when assessing children using an adaptive algorithm above, and because it creates an equal interval scale. Much has been written about the Rasch model (Bond and Fox, 2001; Linacre, 2003; Andrich, 2004), since the original ideas were published a little over 50 years ago.

The full scale

In establishing the validity of the measures for use in this analysis, it is important to build on the previous work mentioned above, and to:

a) Confirm that children's performance on the PIPS assessment can be placed on an equal interval scale. This was done using Rasch[8] measurement techniques. PIPS showed very strong psychometric properties. In particular, the derived performance measures were shown to be extremely reliable person measures with good separation.

b) Check that it is appropriate to report the results on a single scale. A section of Appendix 5 investigates this under the heading "dimensionality" and it confirms that the total score forms a coherent scale. It also shows that the expected sub-scales of *early reading*, *early mathematics*, *vocabulary* and *phonological awareness* can be separated out statistically. It further identifies an interesting subscale which might be termed "symbol identification" comprising letter and number identification.

c) Ensure that the items behave similarly for the different subgroups. In other words, it is important to ensure that items are not biased for or against particular groups[9]. One of the advantages of Rasch Measurement is that if the assessment is working well, the item difficulties should be the same for each sub-group. Rasch essentially creates a ruler which remains constant for all groups of interest. This is in contrast to traditional methods of analysis where the difficulties of items are dependent on the particular group being assessed.

[8] Technically Rasch measurement can be seen as one of a family of models which come under the heading of Item Response Theory but many argue that Rasch measurement is a unique approach that gets to the heart of assessment (fundamental measurement) in a way that other methods do not.

[9] Technically Differential item functioning (DIF) analyses were conducted for age, gender, EAL and deprivation. The details of cut-scores and p values used to identify DIF are given in the appendices.

Item behaviour by age

There was no evidence that the items behaved differently for pupils grouped by age bands. (Four age categories were used: 4 to 4.25[10], 4.25 to 4.5, 4.5 to 4.75 and 4.75 to 5).

Item behaviour by gender

Twelve items showed Differential Item Functioning (DIF) by gender. Four significantly favoured girls; writing, two vocabulary items and the letter Q, (bizarrely) whereas eight of the number identification items were relatively easier for boys. These were not thought to be sufficient to affect the properties of the scale unduly[11].

Item behaviour by EAL

A large number of the items behaved differently for EAL pupils compared to those whose first language was English and those from whom data were missing on this variable. PIPS does allow communication in the child's own language and there are three languages other than English available in the software. However, the vocabulary section assesses English vocabulary acquisition and the phonological awareness section has the same target sounds in all languages. Furthermore, although delivery in different languages is allowed, it is necessary for the administrator to be fluent in that language for the assessment to run smoothly.

Since many of the items were language-based and all involved language, it is hardly surprising to see that the difficulty of items varied, amongst other things, the order in which vocabulary is acquired differs between first and second language learners.

It follows that further analysis of PIPS data to compare EAL children with others at the start of the year is potentially problematic.

Item behaviour by deprivation

The IDACI scores of deprivation of pupils' neighbourhoods were used to create three equal sized groups. Nine items behaved significantly differently by deprivation: one vocabulary item, one Ideas about Reading item, one repeating words item, two letters (including Q; what is it about Q?), one sentence, one item from the Ideas about Maths group and two of the very hard maths items. This was not considered to greatly affect the measurement overall because they represented a small proportion of the total number of items.

---

[10] This group corresponds to the summer born children the very youngest being born in August

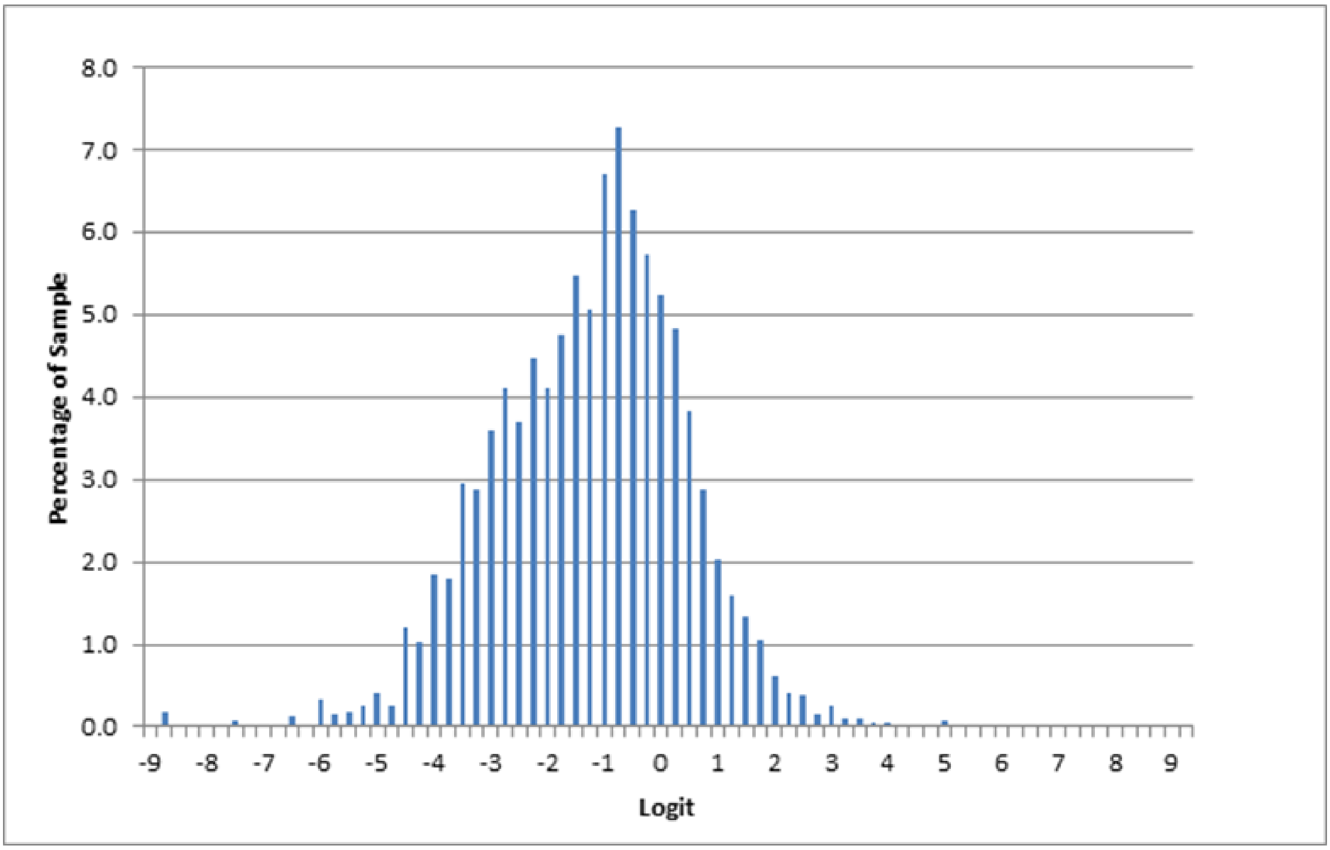
[11] The correlation between the total score with and without the 12 items was 0.999

Item map

The map of the items and the individuals, excluding children with EAL, is shown in Figure 3. The equal-interval scale which is produced is represented by logits. The higher the logit, the more difficult the item or, in the case of a child's ability, the more able the child. The item map gives a very clear idea of what children know and can do when they start school. It also shows a normal distribution of pupils against the general measure which we may call"cognitive development" at the start of school. The detailed percentages of pupils in each bin are given in Figure 4.

Figure 3: Map of items and pupils for the Total measure in England

Figure 4:Distribution of scores on the total measure



The items mapped on the right hand side in Figure 3 show us that the average child starting school in England could easily answer vocabulary items such as “butterfly”, “kite” and “carrots” but that virtually no children could identify a “yacht” or “cosmetics”; those words were outside most children’s vocabularies at that particularly stage. The average (or median) child was likely to know the word “jewellery” as well as “padlock”. In terms of early reading, they were likely to have identified a number of letters such as “T” and about half will know the letter “N” and “J” but they are unable to read sentences, although the most able could identify words such as “dog”, “duck” and “car”.

In terms of *early mathematics*, a high proportion were able to identify single digit numbers such as “6” and “9”, but two digit numbers were beyond most children starting Reception. Similarly the question “What is two more than six?” is beyond the ability of most children at this stage.

In the following sections, the full scale is broken down into four major dimensions: *early mathematics*, *early reading*, *picture vocabulary* and *phonological awareness*. Each is analysed, excluding the EAL children.

Early mathematics

The *early mathematics* items form a single reliable measure with good internal reliability (details are provided in Appendix 2). There is some evidence of additional dimensions in the form of number identification and informal sums/ideas about maths but the scale can be treated as one.

There was little evidence of items behaving differently by children's age, deprivation or gender although girls did find the identification of double digit numbers relatively harder than boys. The map of *early mathematics* items and pupils is displayed in Figure 5 below.

Figure 5: Map of items and pupils for the early mathematics measure in England

```
 5440 PERSON   66 ITEM
-------------------------------------------------------------------------------------

MEASURE    PERSON - MAP - ITEM
               <more>|<rare>
    8                +
                     |  M or L 21 more
                     |T
    7             .  +
                  .  |  M or L 8 more t
                     |  Sums B 15+21
    6                +  Sums B 42-17     Sums B 9-6
                  .  |  M or L 10 less   Sums B 4+11
                     |  M or L 6 less t  Sums B  half of
    5             .  +
                  .  |
                     |
    4             .  +  Sums B  12p ora  Sums B 7+3
                  .  |S Num 3dig1        Sums B 15-4      Sums B Sequence
                  .  |  M or L 3 less t  M or L 3 more t  Num 3dig2        Num 3dig3
    3             .  +
                  .  |  Sums B Sequence
                  .  |  M or L 2 more t  Num 2dig1
    2            .# T+  Num 2dig2        Num 3dig4
                 .#  |  Num 2dig3        Num 3dig5        Sums A rockets
                .##  |
    1          .###  +
              .####  |  Num teen2
           .####### S|  Num teen1        Num teen3        Sums B  half th
    0     .########  +M M or L 1 more t  Sums A pips
       .###########  |  Sums A ice crea  Sums A puppies
      .############  |  Sums B  5p appl
   -1  .###########  +  Sums A bikes     Sums A rabbits
       .########### M|  Sums B  three q
         .#########  |  Num 9
   -2      .#######  +  Num 0            Num 6            Sh hexagon
          .########  |  Num 7            Num 8            Sums A cars
           .#######  |  Count fish here  Count fish ther  IAM least
   -3      .####### S+  Num 5            Sums A balls
            .######  |  IAM shortest     Num 2            Num 3
             .#####  |S Count apples th  Num 1            Num 4            Sh square
   -4          .###  +  Sh triangle
                 .#  |  Count apples he
                 .# T|  IAM most         Sh circle
   -5             .  +  IAM tallest
                  .  |  IAM more
                  .  |
   -6             .  +
                  .  |  IAM smallest     Sh star
                  .  |  IAM biggest
   -7                +
                     |T
                  .  |
   -8             .  +
               <less>|<frequent>
 EACH "#" IS 39. EACH "." IS 1 TO 38
```

The map gives a clear picture of what children know and can do in relation to *early mathematics* when they start school in England. The least able children could use words such as “biggest” correctly but had trouble counting four objects. By contrast, children in the middle of the distribution could identify single digit numbers and were able to respond to questions involving simple informal sums. The more able children could do more complex informal sums and identify two digit numbers, but it was a very unusual child who could do formal sums which needed the knowledge of symbols such as “+”, “-“ and “=” which are generally not known before being taught at school.

Early reading

The *early reading* items form a single reliable measure (details in Appendix 2). There is some evidence of additional dimensions in the form of letter identification as one group, the reading of passages as another and *ideas about reading* as a third but the whole section can be treated as one.

Using the criteria set out earlier there was little evidence of the items in this scale behaving differently by children’s age, deprivation or gender with just one item behaving differently by age, none by gender and none by deprivation.

The map of *early reading* items and pupils is displayed in Figure 6.

Figure 6: Map of items and pupils for the early reading measure in England

```
5440 PERSON  58 ITEM  116 CATS WINSTEPS 3.74.0
------------------------------------------------------------------------------------

MEASURE     PERSON - MAP - ITEM
                <more>|<rare>
    6               .  +
                       |
                       |
                    .  |
    5               .  +  Pass  Walking t
                    .  |
                    .  |T
                    .  |
    4               .  +
                    .  |  Sent  Sentences
                    .  |  Sent  Sentences
                    .  |  Sent  Sentences
    3               .  +  Pass  Cats long  Sent  Sentences  Sent  Sentences  Sent  Sentences  Sent  Story 2
                    .  |  Sent  Story 1
                    .  |
                    .  |S Sent  Sentences
    2              .#  +  W ball           W tree
                   .# T|  IAR full         W flower
                   .#  |
                   .#  |  IAR capital      W car
    1             .##  +  IAR sentence     W duck
                  .##  |  W butterfly      W dog            W house
                 .###  |  Lett Q           W cat            W rabbit
                 .###  |  Lett D           Lett Y
    0           .#### S+M IAR sentence     IAR start        Lett J
               .#####  |  Lett G
                .####  |  Lett H           Lett V           Lett u
               .#####  |  Lett n           Lett t           Lett w
   -1         .######  +  Lett F           Lett K
             .#######  |  Lett A           Lett B           Lett L
              .######  |  Lett c           Lett e           Lett i           Lett p           Lett r           Lett z
             .#######  |  Lett x
   -2         .###### M+  IAR Letter       Lett o
          .##########  |S Lett m
             .#######  |  IAR word         Lett S
              #######  |
   -3     .##########  +
                    .  |  Lett Name Lette
          .##########  |
                 .### S|
   -4                  +
        .############  |
                   .#  |T IAR writing
                       |
   -5                  +
              .######  |
                    .  |
                      T|
   -6                  +  IAR someone
                       |
                  .##  |  IAR someone
                       |
   -7              .#  +
                <less>|<frequent>
EACH "#" IS 36. EACH "." IS 1 TO 35
```

Children generally learn to read after they start school and the map indicates that on entry, almost all children had some knowledge of the *ideas about reading*. The average child could identify the first letter of his or her name and perhaps one or two more letters. The top quartile of children were starting to read simple words and a much smaller proportion could read simple short sentences. Very few were able to read longer texts.

Picture vocabulary

The *vocabulary* items go well together to form a single reliable measure (details in Appendix 2). There was no evidence of additional dimensions.

Figure 7:Map of items and pupils for the vocabulary measure in England

```
5440 PERSON  23 ITEMS
--------------------------------------------------------------------------------------

MEASURE    PERSON - MAP - ITEM
               <more>|<rare>
    7             .  +
                     |
                  .  |  PV cosmetics
                     |
                     |
    6                +  PV yacht
                     |
                     |T
                     |
               .### T|
    5                +
                     |
                     |
                     |
                     |
    4       .#######  +
                     |
                     |  PV microscope
         .######### S|
                     |  PV jewellery
    3             .  +
                     |S PV saxophone
             .#####  |
            .######  |
                     |
    2           .##  +  PV toadstool
          .########  |  PV cash          PV padlock
            .###### M|
                     |
             .#####  |
    1                +
      .############  |
                     |
                  .  |
            .######  |
    0                +M PV pigeon
             .##### S|  PV violin
                  .  |  PV bowl
                .##  |
                .##  |  PV windmill
   -1             .  +
                .##  |
                  .  |  PV knife         PV pan
                 .#  |  PV wasp
                  .  |  PV cherries
   -2             .  +
                 .# T|  PV fork
                     |  PV butterfly
                     |
                  .  |S PV kite          PV turtle
   -3             .  +  PV cupboard
                     |
                     |  PV carrots       PV castle
                     |
                  .  |
   -4            .#  +
               <less>|<frequent>
 EACH "#" IS 56. EACH "." IS 1 TO 55
```

Using the criteria set out earlier there was little evidence of differential behaviour by children's age, deprivation or gender. Just two items showed evidence of behaving differently by deprivation and two different items by gender.

Words such as ‘carrots’ and ‘castle’ were known by almost all children starting school, whereas words like ‘cash’ and ‘padlock’ were known by about half the children. The hardest words were ‘cosmetics’ and ‘yacht’ and were known by very few children starting school.

Phonological awareness

The *phonological awareness* items formed a reliable measure (details in Appendix 2). Although dimensional analysis clearly indicates that the scale could be split into measures of rhyming and repeating words, it was analysed as a single dimension here.

Using the criteria set out earlier there was no evidence of the items behaving differently by children’s age, deprivation or gender.

Figure 8: Map of items and pupils for the phonological awareness measure in England

```
5440 PERSON  17 ITEMS
-------------------------------------------------------------------------------------

MEASURE    PERSON - MAP - ITEM
               <more>|<rare>
    4     .########  +
                     |
                     |
                     |
                     |
                     |
                     |
                    T|
    3    .#########  +
                     |
                     |
                     |
                     |
                     |
         .#########  |
                     |
    2                +
                     |T
        .########## S|
                  .  |
                  .  |
       .###########  |
              .####  |
               .###  |  Rep observatory
    1 .############  +S Rhy dish
          .########  |  Rhy mouse        Rhy pan
          .########  |  Rhy bin
           .#######  |
               .###  |
          .######## M|
          .########  |  Rep frigglejang  Rhy drum         Rhy toes
                .##  |  Rep juxtapose    Rhy cherries     Rhy sun
    0     .########  +M Rhy hat
               .###  |
              .####  |
                .##  |
          .########  |  Rep denalty      Rep riotous
                .##  |  Rep enterprisin
                ###  |
             .##### S|
   -1            .#  +S
                 .#  |
               .###  |
              .####  |
                     |
                  .  |  Rep mantle
            .######  |
                .##  |T
   -2             .  +
                     |
                    T|
                     |
                     |
                  .  |  Rep stop
              .####  |
                  .  |
   -3         .####  +
               <less>|<frequent>
 EACH "#" IS 29. EACH "." IS 1 TO 28
```

The least able children had difficulty repeating words such as ‘mantle’ and ‘stop’ but the average pupils could repeat most words that were presented to them and were starting to get the idea of rhymes. The most able groups could repeat all words presented to them and could correctly pick out all rhyming words.

Differences between groups

The previous section indicated that the measures formed good scales which are comparable across groups, except those for whom English was an additional language. Those children were therefore excluded from the figures below which quantify the differences in ability between boys and girls, the links to age, deprivation and attendance at pre-school.

Before comparing groups, it makes sense to look at the association between the scales. The correlations are presented below.

Table 7: Correlations between the scales for children in England whose first language is English

	Normal Score of Total	Normal Score of Reading	Normal Score of Maths	Normal Score of Picture Vocabulary	Normal Score of Phonological Awareness
Normal Score of Total	1	.882	.898	.641	.694
Normal Score of Reading	.882	1	.734	.460	.493
Normal Score of Maths	.898	.734	1	.508	.549
Normal Score of Picture Vocabulary	.641	.460	.508	1	.476
Normal Score of Phonological Awareness	.694	.493	.549	.476	1

The figure is based on 4,816 cases and all correlations are significant at the 1% level.

The scales all correlated with one another positively at least the 0.48 level which is not surprising, since they all measure general cognitive development. The *early reading* and *early mathematics* measures were most strongly related with a correlation of 0.7. *Picture vocabulary* and *phonological awareness* correlated with *early reading* and *early maths* at around the 0.5 level.

Differences between mean scores are presented as effect sizes and the spread of scores are also recorded since there are clear differences between boys and girls for some measures.

Table 8: Differences between the sexes at the start of the year

Measure	Effect Size (Male-Female)[12]	Standard Deviation Male	Standard Deviation Female
Total	-0.25**	0.85	0.81
Early Reading	-0.26**	1.01	0.95
Early Mathematics	-0.12**	0.87	0.76
Picture Vocabulary	-0.27**	0.87	0.83
Phonological Awareness	-0.23**	0.90	0.89

The Effect sizes (Hedges' d) were calculated using the pooled standard deviation
** significant at the 1% level

The girl's scores were higher, on average, than the boy's scores for the total score, *early reading*, *picture vocabulary* and *phonological awareness* with an effect size of about 0.25. This can be categorised as a "small effect" following Cohen (1969) and who referred to 0.2 as being small, 0.5 as medium and 0.8 as large. For *early mathematics*, the girls were still ahead but only by an effect size of 0.12. Interestingly, the boys' scores were more varied than the girls' scores and this was particularly noticeable for early mathematics.

Differences by age

The figure below shows how the scores correlate with age. These are converted to effect sizes in the last column.

Table 9: Correlations with age at the start of the year

Measure	Correlation with age	Effect Size
Total	0.31**	0.67
Early Reading	0.25**	0.53
Early Mathematics	0.30**	0.65
Picture Vocabulary	0.22**	0.45
Phonological Awareness	0.21**	0.44

** $p<0.01$
Effect size calculated using the formula from Fitz-Gibbon and Morris 1987[13]

[12] Excluding items which behaved differently for males and females (DIF) made a difference to the fourth decimal point for the total score.

Unsurprisingly, all the measures were positively correlated with age and this was most strongly evident for the total score and *early mathematics* where the correlation was about 0.3 corresponding to an effect size of nearly 0.7; a medium to large effect. To get a clearer indication of what this means the regression equation linking the total score to age is given below:

$$\text{Total} = -4.43 + 0.86 * \text{age (in years)}$$

It shows that the total score rises by about 0.07 standard deviation units per month or a difference of 0.86 standard deviation units between the youngest and eldest in a Reception class on average. This corresponds, for example, to the progression from counting a few objects to doing informally presented sums. It also corresponds to the move from knowing a few letters to reading some words.

Difference by deprivation

The correlations of the five measures with deprivation are shown in Table 10.

The strongest link is with the total score and a correlation of -0.2 and indicates that the more deprived had lower scores on average. This is significant in statistical and substantive terms, but is weaker than the correlations that appear for these measures with older children.

Table 10: Correlations with deprivation at the start of the year

Measure	Correlation with IDACI	Effect Size
Total	-0.20^{**}	-0.41
Early Reading	-0.16^{**}	-0.33
Early Mathematics	-0.17^{**}	-0.35
Picture Vocabulary	-0.17^{**}	-0.34
Phonological Awareness	-0.17^{**}	-0.36

$^{**}p<0.01$

Effect size calculated using the formula from Fitz-Gibbon and Morris 1987.

NB1 If EAL children are included, the correlations become stronger, e.g. for the total score it becomes -0.25

NB2 The negative correlations indicate that greater deprivation is associated with lower levels of attainment

---

[13] Based on Hedges and Olkin (1985) The formula is

$$ES = \frac{2r}{\sqrt{(1-r^2}}$$

Differences by pre-school experience

The number of terms at pre-school was recorded by some of the teachers into the contextual data section of the software. These data were provided on an optional basis and so limited samples are available and caution must be exercised in analysing and interpreting the information. Therefore, the breakdown of results and the links to attainment are to be found in Appendix 4 where the conclusion is, that although based on limited data, the data suggests that one or two terms has little impact on the total score at the start of school, but that three or more terms are positively linked to the total score after taking into account age and deprivation. Further analysis does not suggest that that the conclusion varies across deprivation groups.

What levels of personal social and emotional development do children have at the start of Reception in England?

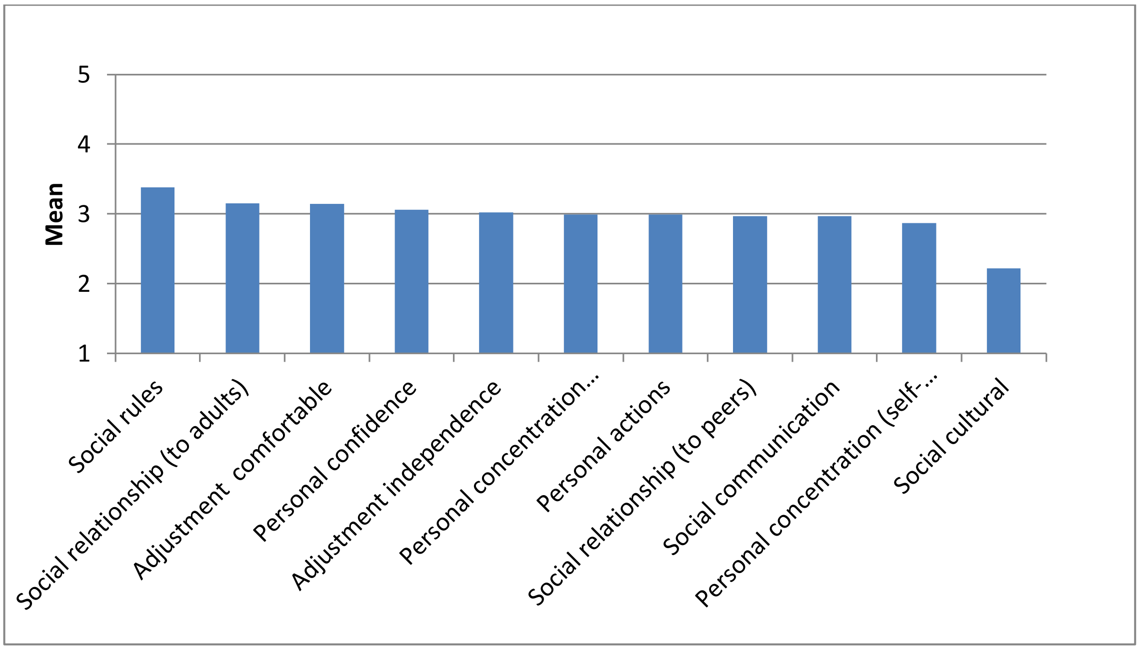
Just over 1,500 pupils were rated by their teachers on the personal, social and emotional development (PSED) items in the PIPS baseline assessment a few weeks after starting school. The assessment was repeated at the end of the school year. Children with EAL were included in this analysis because their development was assessed through teachers' ratings and their assessments should not be dependent on the child's language although the children may differ on that dimension as is investigated below.

The PSED section consists of 11 areas. Ratings for each area were made on a five point scale going from low to high. A full list of items can be found in Appendix 6. The results are presented in Table 11 and Figure 9. It can be seen that, on average, teachers rated children around the middle of the scale although on the *cultural awareness* item the score was an exception to this with the mean closer to two. The spread of scores, the standard deviation, is important because it is large and suggests that a noticeable proportion of children were given the highest rating of five and some were given the lowest rating of one for each item.

Table 11: PSED ratings

	N	Mean	Standard Deviation
Social rules	1,565	3.38	0.85
Social relationship (to adults)	1,565	3.15	0.85
Adjustment comfortable	1,566	3.14	0.95
Personal confidence	1,565	3.06	0.91
Adjustment independence	1,565	3.02	0.86
Personal concentration (teacher-directed)	1,565	2.99	0.88
Personal actions	1,565	2.99	0.87
Social relationship (to peers)	1,565	2.97	0.76
Social communication	1,565	2.97	1.04
Personal concentration (self-directed)	1,565	2.87	0.88
Social cultural awareness	1,565	2.22	0.85

Figure 9: Children's PSED at the start of school in England



The PSED ratings were further analysed by the gender of the child, their age, the IDACI score derived from the home postcode deprivation, attendance at pre-school and whether the child's first language was English. Table 12 shows the differences between groups, expressed as effect sizes. For all the PSED items, girls were, on average, rated at a higher level than the boys. The advantage was as high as nearly 0.5 of standard deviation, or just less than half a point, and all comparisons were statistically significant except for the items relating to *feeling comfortable on arrival at school*. One must be wary when interpreting this kind of information because it is based on teachers' judgements, which may vary from teacher to teacher and may be biased.

The older the child on starting school, the higher the ratings tend to be on each item. The effect sizes are modest but clear; the older children were seen to concentrate more, to feel more comfortable, to communicate better, to have better relationships and so on. There was a fairly constant effect across all items.

The link to deprivation was weak and much smaller than the links to gender and age. It was non-significant except for *concentration on self-directed activities* and for *rules* where the more deprived the child's home background, the lower their rating on the *concentration* items and the less they were seen as responding positively to *rules*. *Concentration* is important because it relates positively to later academic performance, just as inattention is a predictor of less progress in later years (Merrell and Tymms, 2001; Merrell and Bailey, 2012; Pingault, Tremblay, Vitaro et al, 2011).

Attendance at nursery was investigated but, as noted earlier, the data were collected optionally and may not provide a representative picture. The analysis is shown in Appendix 4 where positive links are found between attendance at nursery and all items. They were found to be statistically significant for seven of the items.

Table 12: Links between PSED and gender, age, deprivation and EAL in effect sizes

Measure	Male-female	Age	Deprivation	1st Language English - EAL
Social rules	-0.46**	0.25**	-0.18**	0.03
Social relationship (to adults)	-0.28**	0.35**	-0.02	0.23**
Adjustment comfortable	-0.10	0.36**	0.01	-0.05
Personal confidence	-0.16**	0.45**	-0.04	0.22**
Adjustment independence	-0.29**	0.42**	0.02	-0.03
Personal concentration (teacher-directed)	-0.43**	0.38**	-0.13	0.06
Personal actions	-0.46**	0.28**	-0.08	0.08
Social relationship (to peers)	-0.35**	0.30**	-0.06	0.26**
Social communication	-0.30**	0.33**	0.03	0.44**
Personal concentration (self-directed)	-0.47**	0.44**	-0.18**	0.05
Social cultural awareness	-0.26**	0.23**	0.06	0.15

All figures are based on 1,565 (+-2) cases except the nursery attendance data which is based on 661 cases.
** p<0.01

The last column in Table 12 compares the children whose first language was English with those whose first language was identified as not English and there were just a few differences. The children with EAL were rated as less *confident* and not so adept in *relating to both other children and to adults* when compared to children whose first language is English. Not surprisingly, the biggest differences were seen in their *communication* skills ratings.

Progress in cognitive development and PSED between the start and end of Reception year in England

Cognitive development (excluding children with EAL)

The gains in cognitive development during the first year of school are set out in Table 13. The growth in this first year is massive with an effect size of 2.44 for the total score. Effect sizes in education are usually less than one (Hattie,1999; Hill, Bloom, Black et al, 2008). As noted earlier, Cohen (1969) suggests that an effect size of 0.8 to be considered as large. The effect size reported here (2.44) is amongst the largest reported in education research and in agreement with Hill, Bloom, Black et al (2008), is probably larger than any effect size in any other year at school. One would expect an effect size of around 0.3 per year during much of the primary stage and then less than that as the children progress through secondary school[14]. Of course, during the first year at school children start to acquire basic skills and it is not unreasonable that they will make the longest strides early on; but they really do seem to be very long strides and it may be worth recalling that cognitive growth curves often have an elongated S shape with slow progress at the start followed by rapid gains, which then flatten off (Demetriou and Raftopoulos, 2005).

Part of the gain is due to maturation, as the children were three quarters of a year older when they were assessed for the second time. The third column in the figure sets out the gain calculated to be due to maturation alone, and in the case of the total score this amounts to an effect size of 0.48[15]. It follows that the effect size due to schooling is 1.96, which is still a very large gain.

The figure also sets out the gains for *early reading*, *early mathematics*, *picture vocabulary* and *phonological awareness* separately. The major progress was in *early reading* and *early mathematics* but very large gains were also made in *vocabulary* and *phonological awareness.*

[14] This is based on the expected level of progress in terms of levels per year and the standard deviations at the ends of Key Stages.

[15] There is a linear relationship between PIPS score and age. This was used to estimate the gain that would be expected by maturation alone.

Table 13: Gains in cognitive development for England (excluding EAL)

Measure	Effect Size overall	Effect Size by maturation	Effect Size due to schooling
Total	2.44	0.48	1.96
Early Reading	2.49	0.73	1.76
Early Mathematics	2.50	0.72	1.78
Picture Vocabulary	1.47	0.57	0.90
Phonological Awareness	1.49	0.60	0.89

Personal, social and emotional development

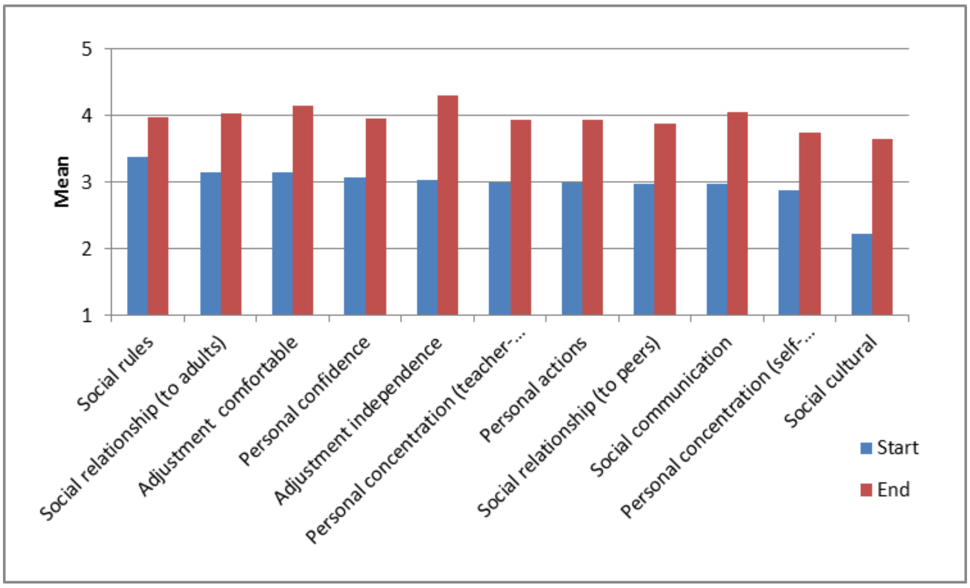
Teachers rated their pupils on the same PSED items at the start and the end of the Reception year and very large gains were seen, typically around one standard deviation, or one point on the rating scale, although some were larger than that, as reported in Table 14 and Figure 10. Some of the gain is due to maturation and it was noted earlier that age was clearly related to all the items. As with the cognitive development measures, the gain explained by maturation alone was calculated and it is reported in the third column of the figure. This was used to generate the fourth column which shows the residual gain associated with schooling.

Clearly, some children were hitting the top of the scale and therefore the estimates are probably understated. Nevertheless the gains are substantial. All of the effect sizes are at least 0.25 and most are much larger, especially the one for cultural awareness, which was 1.13.

Table 14: Gains in PSED measures for England (including EAL)

PSED Measure	Effect Size Overall	Effect Size by Maturation	Effect Size due to Schooling
Social rules	0.67	0.33	0.34
Social relationship (to adults)	0.98	0.48	0.50
Adjustment comfortable	1.04	0.49	0.64
Personal confidence	0.96	0.60	0.36
Adjustment independence	1.42	0.60	0.82
Personal concentration (teacher-directed)	1.01	0.49	0.52
Personal actions	1.03	0.36	0.67
Social relationship (to peers)	1.07	0.38	0.69
Social communication	1.07	0.48	0.59
Personal concentration (self-directed)	0.90	0.53	0.27
Social cultural awareness	1.45	0.29	1.13

Figure 10: PSED at the start and end of the Reception year in England



Comparing jurisdictions

The next part of the report compares the cognitive development of children starting school in different jurisdictions, taking into account the age of those children. Before making comparisons, it is important to ensure that the items from the assessments behave similarly across jurisdictions and Rasch measurement techniques were used to check this.

Vocabulary was not compared because language varies so much across cultures and the items had different orders of difficulty in different jurisdictions. For example, an analysis by items revealed that the item 'wasp' from the vocabulary section was particularly difficult for children in Australia compared with the other jurisdictions whereas the item 'toadstool' was particularly difficult in England. After removing all the items that behaved differently across the jurisdictions, just eight items remained, which was too few to form a reliable scale for making comparisons.

The *phonological awareness* scale has been adjusted in Australia, where a bespoke section had been introduced, leaving too few items in common to be able to make comparisons between jurisdictions.

For both *early mathematics* and for *early reading*, there were sufficient items to be able to proceed.

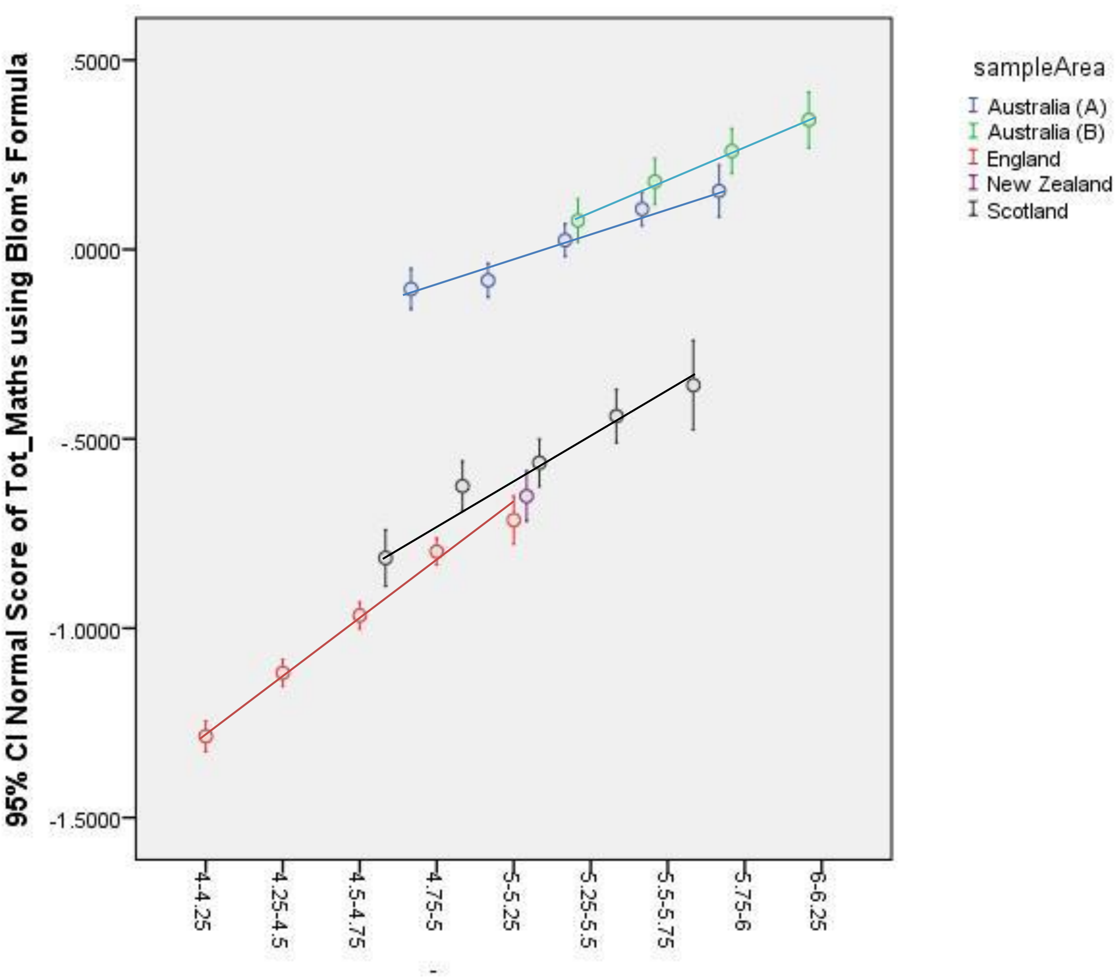
Early mathematics at the start of the year

Forty of the *early mathematics* items behaved differently across jurisdictions and they were excluded from further analysis, leaving a group of items which were comparable. These items were used to create a measure of *early mathematics* for each child which was comparable across jurisdictions.

To control for children's varying starting ages across countries, the age at the time of assessment for each child was calculated from the dates of birth and these were grouped into categories, split into quarter year slices. The youngest category included children from age four to four and a quarter years, and the oldest included children from age six to six and a quarter years. Figure 11 is a plot of the average ability score for each of the age categories, restricted to groups where there were at least 200 children.

As expected, there was a positive relationship between ability scores and age. The children starting school in England were the youngest of the jurisdictions included in the analysis. The ability scores rose quite noticeably by age, as did the scores for Scotland. Both of the Australian groups had high scores for their age and the relationship to age is rather flatter than would be expected by comparison with the English and Scottish data. The children in New Zealand start school just after their fifth birthday and were in line with the results from England and Scotland.

Figure 11: Early mathematics at the start of school (without EAL)

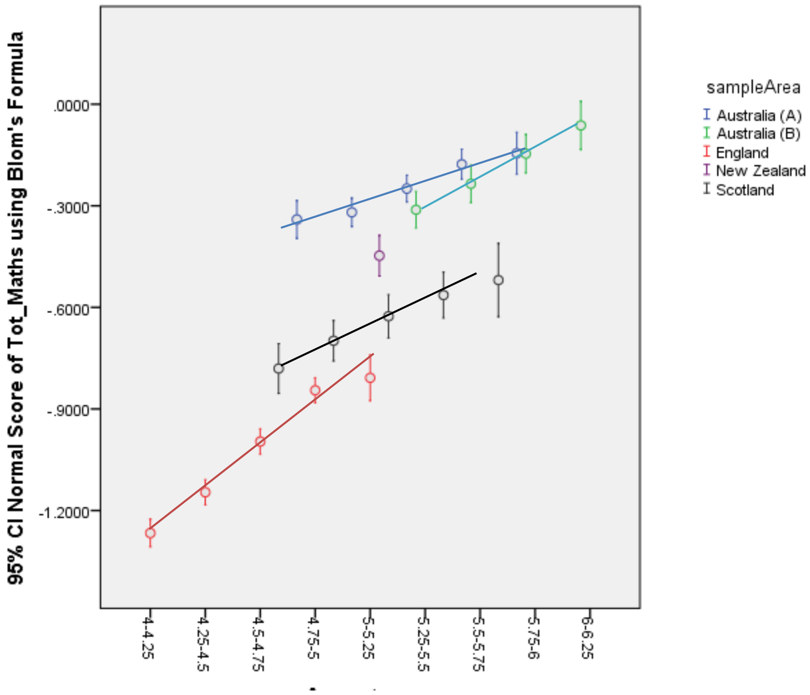


Early reading

Following an analysis of the behaviour of items across jurisdictions, the items which varied in relative difficulty across jurisdictions were removed to leave a scale of 44 items. Some of the easier items in the *ideas about reading* section showed some variation in behaviour between jurisdictions but they were retained in the scale to ensure that the lower end of the scale was sufficiently represented.

Figure 12, which compares *early reading* scores by age category, shows similar patterns for England and Scotland as were seen for *early mathematics*, with an increase in ability by age, although the children in Scotland were slightly ahead of those in England and the line was a little flatter. The two Australian groups were ahead of the children in the other jurisdictions, age for age, and the children in New Zealand were positioned between the two.

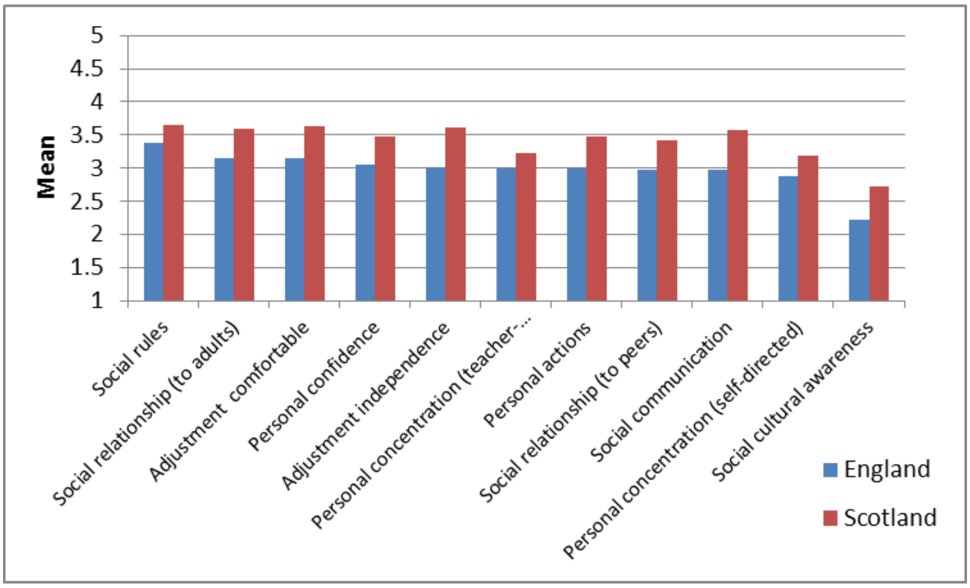
Figure 12: Early reading at the start of school (without EAL)



Personal, social and emotional development

There was sufficient data to make comparisons for England and Scotland but not for other jurisdictions. Figure 13 shows the results. Although the children from Scotland scored higher than the children from England, they were, on average, six months older and we have already noted that the PSED scores increase with age. However, further investigation using the General Linear Model and controlling for age showed that the ratings for children were still higher for Scotland than for England by between a tenth and a half a point on the rating scale. These differences were significant at the 1% level except for the two ratings of concentration.

Figure 13: PSED at the start of school (with EAL) in England and Scotland



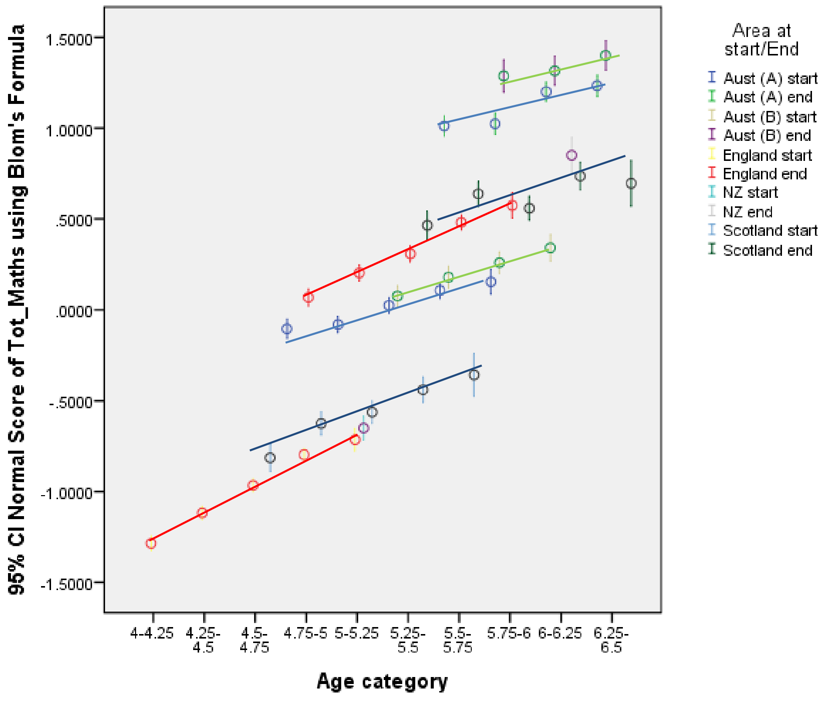
Progress across jurisdictions

The progress during the first year at school is analysed and presented in two ways. First the mean pupil ability scores were compared at the start and the end of the year, and this progress is reported graphically. Then multi-level models were run with two levels; pupils and schools with jurisdictions identified by dummy variables. The *early reading* and *early mathematics* scales used in the previous section were analysed.

Early mathematics across jurisdictions

Figure 14 shows the pupils' ability scores by age category at the start and the end of the year for the five jurisdictions. The graphical representation of the data for England indicated a very large impact of schooling and this was paralleled by the data from Scotland. The other jurisdictions exhibited similar rises, as schooling in all areas appeared to have similar impacts. The multi-level models, which are set out later, give a more accurate picture of changes and quantify, for example, the different amount of time between assessments of the New Zealand children compared to the other jurisdictions. They were assessed a year later whereas in the other areas they were assessed at the start and end of the same school year.

Figure 14: Start and end of year scores for early mathematics (no EAL)

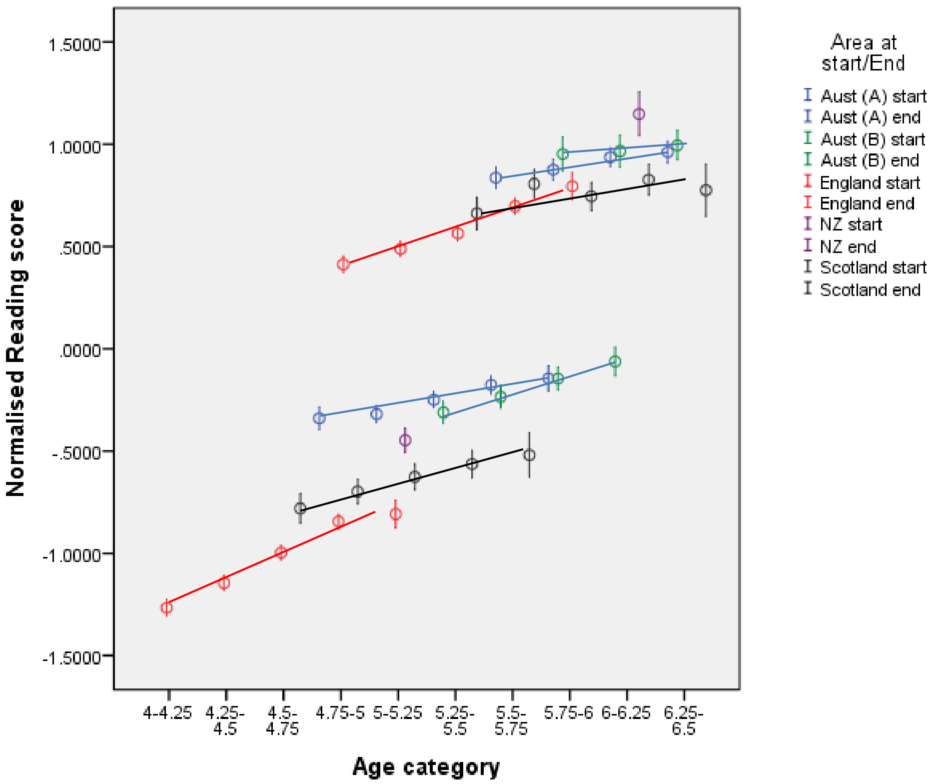


NB There are fewer points for Australia A and B at the end of the year compared with the start because some cases were lost during the year and points were only plotted on the chart where there were at least 200 cases.

Early reading across jurisdictions

The patterns for *early reading* were different from those seen for *early mathematics*. Although all jurisdictions showed large gains, the children in England seemed to make more progress, closely followed by the children in Scotland and New Zealand. By the end of the year, although all the jurisdictions have close scores for each age categories, there are still differences. New Zealand has the highest score of all.

Figure 15: Start and end of year scores for early reading (no EAL)



NB There are fewer points for Australia A and B at the end of the year compared with the start because some cases were lost during the year and points were only plotted on the chart where there were at least 200 cases.

Figure 15 could be interpreted as all groups coalescing at a similar high *early reading* score, suggesting a ceiling on the assessment. But further investigation showed that this was not a valid interpretation as the reading scale went much higher than the hypothesised ceiling.

Multilevel models

The multi-level models seek to partition the variance of the end of year scores in *early mathematics* and *early reading* between pupils and schools, and account for that variance by changes to the model. Ideally, one would include good measures of home background but they were not available in the present data because we did not collect comparable socio-economic status measures across jurisdictions. Nevertheless, the key prior attainment measures were available. The models are set out for *early mathematics* and then for *early reading*, which can be related back to the charts in Figures 13 and 14 that display almost the same attainment data used in the models. Unlike the previous analyses of the cognitive measures, the children with EAL were included. This group represented 9.3% of the 19,228 cases studied. A dummy variable was used to separate them from the other children.

Early mathematics

The models are set out in Table 15 below. They build up in complexity from the so-called null model through to Model 3. In each case, more variables were added to the model, thus reducing the variance. In the null model, the Constant (0.59) is the mean for the whole sample and the figure in parenthesis (0.02) shows the standard error on that mean. The end of year mathematics variance was partitioned between pupils (0.20) and schools (0.40); in other words, a third of the variance was found at the school level. In Model 1, dummy variables were used to identify jurisdictions in comparison to England, which had the lowest scores with the youngest children. Just above England came New Zealand and Scotland, which both had a score of 0.37, and above them came the two Australian jurisdictions with Australian B having the highest score of 1.10. The school level variance dropped by just over half, but the pupil variance did not change, leaving around a fifth of the variance at school level.

In Model 2, the key explanatory variables from the start of school were introduced. Since early mathematics is the outcome, it is not surprising to see that the early mathematics score at the start of school was the most important variable with a coefficient of 0.68. Early reading was also important, whereas age had a small positive coefficient. Children with EAL had a positive coefficient, indicating that they made more progress than children for whom English was their first language[16]. The coefficients for the different jurisdictions were very similar and none differed significantly from England. In other words, the progress measures for each jurisdiction were essentially identical. By including the various variables in the model, the pupil level variance dropped by about 60% and a fifth came off the school level variance. In Model 2, a third of the variance was associated with school membership. This is a very high proportion of variance suggesting the school attended made a very large difference to the progress that children made in early mathematics[17]. By contrast, the jurisdiction in which a child was educated made almost no difference.

Finally, in Model 3, the mean school mathematics score at the start of the year was included. It had a negative and significant coefficient, indicating a tendency for the schools with the highest average score at the start of the year to make the least progress. Generally, Model 3 was similar to Model 2, except that Australia B now had a higher and significant coefficient. This suggests that the Australia B schools with their high-scoring children at the start of the year were making better progress than other jurisdictions after taking into account the general finding that schools with higher intakes tend to make less progress.

[16] If the slope is linked to EAL then this positive coefficient becomes non-significant and the slope for the first measure of early maths is less positive for the EAL children

[17] Typically school effectives students report the proposition of variance associated with school membership in the range 10-30% (Guldemond and Bosker, 2009)

Table 15: Multi-level models with early mathematics as the outcome

	Null	Model 1	Model 2	Model 3
Fixed				
Constant	0.59 (0.02)*	0.30 (0.02)*	0.06 (0.05)	0.93 (0.06)
Australia A		0.77 (0.04)*	-0.06 (0.04)	0.02 (0.05)
Australia B		1.10 (0.05)*	0.06 (0.04)	0.18 (0.06)*
New Zealand		0.37 (0.10)*	-0.08 (0.09)	-0.05 (0.09)
Scotland		0.37 (0.03)*	-0.03 (0.02)	0.02 (0.03)
Age			0.03 (0.01)*	0.03 (0.01)*
EAL			0.08 (0.01)*	0.07 (0.01)*
Start Reading			0.16 (0.01)*	0.16 (0.01)*
Start Mathematics			0.68 (0.01)*	0.69 (0.10)*
School av mathematics				-0.09 (0.03)*
Random				
School level variance	0.20 (0.01)*	0.098 (0.006)*	0.080 (0.005)*	0.079 (0.005)*
Pupil level variance	0.40 (0.004)*	0.40 (0.004)*	0.162 (0.002)*	0.163 (0.002)*
% at School level	33.3	19.7	33.0	32.6
% Drop at school level		55.5	20.0	1.2
% Drop at pupil level		0	59.5	-0.6

Figures in parentheses are standard errors ***** p<0.05

Early reading

In the null model for early reading, nearly 20% of the variance was ascribed to the schools. In Model 1, the jurisdictions are identified by dummy variables and England, with the youngest children, has the lowest scores with Scotland being slightly ahead, Australia A was slightly further ahead, New Zealand further ahead and Australia B just ahead of New Zealand. By introducing the dummy variables for jurisdictions, the school level variance dropped by almost a quarter and the pupil level variance dropped by a very small amount. The proportion of variance associated with schools was about 15%.

In Model 2, as for early mathematics, the other measures were introduced into the model and now it is clear that the progress made in England was more than in any other jurisdiction although not significantly more than New Zealand. Scotland made a little less progress whereas England made a lot more progress than the two Australian jurisdictions, which initially had started at a higher level. All the jurisdictions finished at a similar level. Perhaps surprisingly, the age coefficient was negative which meant that the older children made less progress than younger children on average. This can be seen inTable 16. The children with

EAL were, on average, making more progress than others[18]. Early Reading ability at the start of school was the most important variable in the model. After the addition of measures in Model 2, the pupil level variance dropped by over a half and the school level variance by about 14%; it leaves 25% of the variance linked to schools which is a very high proportion in comparison with other school based studies.

In Model 3, the school average reading score at the start of the year was introduced. It had a statistically significant negative coefficient but with only a little impact on other coefficients, however, it did impact on Scotland's coefficient, removing its statistical significance.

Table 16: Multi-level models with early reading as the outcome

	Null	Model 1	Model 2	Model 3
Fixed				
Constant	0.73 (0.01)*	0.56 (0.02)*	1.60 (0.5)*	1.44 (0.06)*
Australia A		0.29 (0.04)*	-0.39 (0.03)*	-0.27 (0.04)*
Australia B		0.42 (0.04)*	-0.38 (0.02)*	-0.24 (0.04)*
New Zealand		0.40 (0.08)*	-0.09 (0.08)	0.02 (0.08)
Scotland		0.23 (0.02)*	-0.09 (0.02)*	-0.03 (0.02)
Age			-0.04 (0.1)*	-0.04 (0.01)*
EAL			0.14 (0.01)*	0.14 (0.01)*
Start Reading			0.50 (0.01)*	0.51 (0.01)*
Start Mathematics			0.32 (0.01)*	0.32 (0.01)*
School av reading				-0.17 (0.02)*
Random				
School	0.084 (0.006)*	0.064 (0.004)*	0.055 (0.003)*	0.052 (0.003*)
Pupil	0.375 (0.004)*	0.365 (0.004)*	0.164 (0.002)*	0.163 (0.002)*
% at School level	18	14.9	25.1	24.2
% Drop at school l		23.8	14.0	1.2
% Drop at pupil l		2.7	55.1	0.1

***** p<0.05

[18] As with mathematics the slope associated for the first best predictor (Early Reading in this case) was significantly more shallow and the intercept for EAL became non-significant.

Is it possible to have an international assessment for children starting school?

The case for creating an international comparative study of the sort that exists for later primary and secondary years (e.g. TIMSS, PIRLS and PISA) has been made in a number of places, and broadly welcomed in principle by policy makers in national governments and international agencies. Not only would such a study help to contextualise and interpret these later assessments, but it would help to shed light, in a way that no other type of study could, on some important policy issues, such as 'how effective are different types of early years programmes?', 'how important is it to have effective teachers in the first year of school?', 'how do different factors influence children's learning?', 'what are the best policies for long term effectiveness?'

This study has not addressed these policy questions directly; rather, it has sought to examine the feasibility of such a study using existing data. This final section therefore builds on the findings in the report to consider whether or not it is possible to compare the development of children starting school across different countries.

There are two major issues concerning comparability. The first is that children start school at different ages in different parts of the world and, at that young age, cognitive progress is very rapid, thus challenging the possibility of comparability. The second is that children in different countries grow up in different cultures, speak different languages and often learn in different ways with different materials. Given these issues, is it possible to make valid comparisons?

The analyses carried out in this paper started by looking at the data from children at the start of school. With its great diversity of content and range of difficulty, the PIPS On-entry Baseline assessment is able to accommodate a very wide range of children. It was not found to be appropriate to include children for whom English was an Additional Language (EAL) for several of the analyses and this point is touched on below. However, the fact that the assessment could accommodate very slow learners at the age of just four years and a day up to strong learners at the age of six and half at the end of their first year of school, and that the data can all be put onto the same scale for *early mathematics* and *early reading* is very promising. This range represents a considerable variation in cognitive development, more or less covering the full range of children starting school around the world. Further, these data can all be collected in around 20 minutes per child. *Early reading* and *early mathematics* can form a single linear scale, using Rasch measurement, and this has been shown to be valid for both sexes, across the full socioeconomic range and across five English-speaking jurisdictions.

The second issue relates to the language and the context of the child. The evidence that we have presented for the differential item functioning in England for children with EAL is not relevant, because these children were being assessed in an English context as opposed to children in a different jurisdiction speaking in the language of that country. It is interesting to note that for the *early mathematics* section the evidence is that there is little difference in how items perform for different groups of children, and it would be quite possible to create a meaningful scale which incorporated children with EAL. It was not possible to follow this up as part of the report because of time constraints.

This analysis highlights some problems with an assessment of *early reading*, particularly regarding vocabulary and phonological awareness. Although not part of this study, other work has shown that PIPS can be adapted quite successfully to other languages and cultures. To date, versions have been created and used in Dutch (where it is known as OBIS), German (where it is known as FIPS), Russian, Spanish, French, Slovenian and Chinese (both Cantonese and Mandarin), Afrikaans and Sepedi (another Southern African language).

The mathematics items do not generally present any problems when translated, although some visual material such as coins need to be changed, and some of the later and/or more able starters (from our research in Russia and China) required the development of a range of harder items to avoid a ceiling effect in the assessment.

The reading items generally prove more challenging to adapt, particularly for languages whose syntax does not bear a very close resemblance to English, however, in those languages where the most research has been carried out to date (German, Dutch and Russian) it has been possible to develop versions of the items which are rated as equivalent by professional linguists and early years educators competent in both languages, and which work well when used with experimental cohorts of children.

The *vocabulary* and *phonological awareness* scales have proved the most challenging in terms of generating equivalent versions for the different languages and cultures, and this lends support to the idea that these two scales should not be used for international comparisons, but adapted for use internally by policy makers within individual countries as these are important features widely regarded as fundamental to children's language acquisition and developing literacy skills.

In conclusion, we believe that it is both possible and desirable to have an international study of children's cognitive and non-cognitive development starting school, and of their progress during the first school year. The research and policy questions are too important to leave unexplored. As with any research, such a study is unlikely to give definitive answers to questions such as what age children should start school, but, in providing an important set of international benchmarks and very useful policy-relevant information within countries, which is not available to many at the present time, it will certainly provide some very important and additional information and help policy makers evaluate existing policies in individual countries.

References

Andrich, D. (2004). Controversy and the Rasch Model: A Characteristic of Incompatible Paradigms? *Medical Care, 42*(1 suppl), I-7 to I-16.

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

CEM (2001). *Performance Indicators in Primary Schools: Technical Report.*

Cohen, J. (1969). *Statistical Power Analysis for the Behavioral Sciences*. NY: Academic Press.

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

Demetriou, A., and Raftopoulos, A. (Eds.) (2005). *Cognitive developmental change: Models, methods, and measurement*. Cambridge: Cambridge University Press.

Department of Education (2001). *National Pupil Database Spring Census 2012.* Unpublished raw data.

Fitz-Gibbon, C. T. and Morris, L. (1987). *How to analyze data*. Beverly Hills, CA: Sage Publications.

Guldemond, H. and Bosker, R. (2009). School effects on students' progress – a dynamic perspective. *School Effectiveness and School Improvement,* 20(2), 255-268.

Hattie, J. (1999). *Influences on student learning: Inaugural Lecture*. Auckland, New Zealand: University of Auckland.

Hill, C. J., Bloom, H. S., Black, A. R. and Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives,* 2 (3), 172-177.

IDACI scores (2010). Pages accessed 10.01.14: http://webarchive.nationalarchives.gov.uk/20120919132719/http://www.communities.gov.uk/communities/research/indicesdeprivation/deprivation10/

Linacre, J. M. (2003). Data Variance: Explained, Modeled and Empirical. *Rasch Measurement Transactions,* 17(3), 942-943.

Merrell, C. and Bailey, K. (2012). Predicting Achievement in the Early Years: How Influential is Personal, Social and Emotional Development? *Online Educational Research Journal.* http://www.oerj.org/View?action=viewPaper&paper=55.

Merrell, C., Styles, I., Jones, P., Tymms, P and Wildy, H. (2013). Cross-country Comparisons of Inattentive, Hyperactive and Impulsive Behaviour in School-Based Samples of Young Children. *International Research in Early Childhood Education,* 4 (1).

Merrell, C. and Tymms, P. (2001). Inattention, hyperactivity and impulsiveness: Their impact on academic achievement and progress. *British Journal of Educational Psychology*, 71, 43-56.

New Zealand Education Counts (2013). Downloaded 10.01.13: http://www.educationcounts.govt.New Zealand/directories/list-of-New Zealand-schools

New Zealand decile (2013). Downloaded 10.01.14: http://www.minedu.govt.New Zealand/New Zealand/EducationPolicies/Schools/SchoolOperations/Resourcing/OperationalFunding/Deciles/HowTheDecileIsCalculated.aspx

Pingault, J. B., Tremblay, R. E., Vitaro, F., Carbonneau, R., Genolini, C., Falissard, B., and Côté, S. M. (2011). Childhood trajectories of inattention and hyperactivity and prediction of educational attainment in early adulthood: A 16-year longitudinal population-based study. *American Journal of Psychiatry*, 168(11), 1164-1170.

Sayal, K., Owen, V., White, K., Merrell, C., Tymms, P. and Taylor, E. (2010). Impact of Early School-Based Screening and Intervention Programs for ADHD on Children's Outcomes and Access to Services. *Archives of Pediatrics & Adolescent Medicine.* 164(5), 462-469.

Scottish Educational Analytical Services Division (2013). Unpublished raw data.

SIMD (2013). Pages accessed 10.01.14: http://www.scotland.gov.uk/Topics/Statistics/SIMD/BackgroundMethodology

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

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

Tymms, P., Brien, D., Merrell, C., Collins, J. and Jones, P. (2003). Young deaf children and the prediction of reading and mathematics. *Journal of Early Childhood Research*, 1(2), 197-212.

Tymms, P., Jones, P., Albone, S., and Henderson, B. (2009). The first seven years at school. *Educational Assessment and Evaluation Accountability,* 21, 67-80.

Tymms, P., Merrell, C., Henderson, B., Albone, S., and Jones, P. (2012). Learning Difficulties in the Primary School Years: Predictability from On-Entry Baseline Assessment. *Online Educational Research Journal*.

Appendix 1: The Samples

Establishing representative samples of PIPS data for England, Scotland, New Zealand and two territories/states in Australia

England

Because schools can choose whether or not to use PIPS, it is necessary to establish first whether the range of pupils' backgrounds is representative of children starting state schools in England and, if it is not, whether a representative sample can be established.

Schools complete background data for each pupil within the assessment, when they start school in Reception aged four. This includes forename, surname, gender and date of birth. These items were used to match the data recorded in the PIPS database to the National Pupil Database held by the Department of Education to allocate an Income Deprivation Affecting Children Index (IDACI, 2010) Score to individuals.

The Spring Census 2012 was linked to the PIPS data for the academic year 2011/2012, which is the cohort analysed within this report.

The IDACI score, by decile, for the home postcode for all state pupils attending Reception class in England 2012 matched to the Pupil Spring Census Dataset (Department for Education, 2013) is shown in Table 17 below.

Table 17: Number of pupils in each decile in Reception in England 2012 (Population 609,983)

Decile	1	2	3	4	5	6	7	8	9	10	Total
Number	86,565	76,002	67,287	62,047	57,870	53,949	53,597	52,741	51,263	48,662	609,983
Proportion	0.14	0.12	0.11	0.10	0.09	0.09	0.09	0.09	0.08	0.08	

This can be compared with Table 18 which shows the distribution of the Pupils' IDACI scores for the home postcode put into the same deciles from the PIPS dataset assessment for autumn 2012 that were matched to the Spring Census data 2012 for Reception children by the Department of Education National Pupil Database.

Table 18: Number of pupils in each decile in Reception in PIPS sample England 2012 (Sample 8,746)

Decile	1	2	3	4	5	6	7	8	9	10	Total
Number	991	1,105	1,093	966	897	684	654	735	794	827	8,746
Proportion	0.11	0.13	0.12	0.11	0.10	0.08	0.07	0.08	0.09	0.09	

The proportions shown in Table 19 are fairly similar, but in order to equate them, an appropriate proportion of pupils were randomly sampled in each decile and to give a representative sample of the National Dataset. The sample was reduced to 6,986 pupils and this sample was used in the subsequent analyses.

Table 19: Number of pupils in each decile in Reception in the PIPS sample for England 2012 matched to National Distribution (Sample 6,986)

Decile	1	2	3	4	5	6	7	8	9	10	Total
Number	991	870	771	711	663	618	614	604	587	557	6,986
Proportion	0.14	0.12	0.11	0.10	0.09	0.09	0.09	0.09	0.08	0.08	

Scotland

The data from Scotland were treated in a similar way although the measure of deprivation was different and the links to that measure were different. As in England, schools completed background data for each pupil within PIPS when they started school in Primary 1 aged between four and a half and five and a half. This includes home postcode and Scottish Candidate Number (SCN). Either of these measures can be used to allocate an individual to one of the deciles linked to the Scottish Index of Multiple Deprivation (SIMD). However, the assessment can be carried out without these details being completed and for the assessments administered in autumn 2012, the following information was available:

Number of pupils assessed: 20,775
Number of valid pupil postcodes: 6,519
Number of valid Scottish candidate IDs: 9,991
Number of pupils with either a valid postcode or SCID or both: 12,123

Table 20 shows the distribution of pupils in Primary 1 by Scottish Index of Multiple Deprivation (SIMD 2013) decile from data supplied by the Educational Analytical Services Division. 335 pupils with unknown data zone are not included.

Table 20: Number of pupils in each decile in Primary 1 in 2012 (Population 56,349)

Decile	1	2	3	4	5	6	7	8	9	10	Total
Number	6,841	6,146	5,568	5,593	5,456	5,417	5,736	5,567	5,440	4,585	56,349
Proportion	0.12	0.11	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.08	

This distribution can be compared with the distribution of pupils in the PIPS dataset based only on postcode information.

Table 21: Number of pupils in each decile from PIPS in 2012 based on postcodes (Sample 6,519)

Decile	1	2	3	4	5	6	7	8	9	10	Total
Number	632	578	706	729	599	687	737	734	683	434	6,519
Proportion	0.10	0.09	0.11	0.11	0.09	0.11	0.11	0.11	0.11	0.07	

The distribution of the number of pupils in the PIPS data set for whom home postcodes were available was fairly even across the deciles. The proportions were close to the full dataset shown in Table 22 with slightly fewer than expected for some groups (deciles 1, 2 and 10).

Table 22: Number of pupils in each decile from PIPS in 2012 based on SCNs (Sample 9,911)

Decile	1	2	3	4	5	6	7	8	9	10	Total
Number	1,015	1,227	1,064	1,194	1,024	914	911	901	1,136	525	9,911
Proportion	0.10	0.12	0.11	0.12	0.10	0.09	0.09	0.09	0.12	0.05	

The distribution of the number of pupils in the PIPS data set for whom SCNs were available was also fairly even across the deciles. The proportions were close to the full dataset shown in Table 23 with slightly fewer than expected for some groups (deciles 1, 6, 7 and 10)

For 4,275 pupils, deciles were available which were derived from a postcode and SCN. The vast majority of these deciles were identical but in some cases they were not. The two sets of deciles were combined and, where there was conflict, the decile derived from the SCN was preferred. The resultant deciles derived from both postcodes and SCN are shown below inTable 23.

Table 23: Number of pupils in each decile from PIPS in 2012 based on SCNs and postcodes (Sample 12,123)

Decile	1	2	3	4	5	6	7	8	9	10	Total
Number	1,147	1,372	1,264	1,437	1,240	1,170	1,222	1,189	1,373	709	12,123
Proportion	0.10	0.11	0.10	0.12	0.10	0.10	0.10	0.10	0.11	0.06	

A proportional random sample of pupils in each decile from PIPS data was selected to give a representative sample of the National Dataset as supplied by the Scottish Government Education Analytical Services. This is reported inTable 24. This sample was used for further investigation.

Table 24: Number of pupils in each decile in Primary 1 in PIPS sample Scotland 2012 matched to National Distribution (Sample 6,627)

Decile	1	2	3	4	5	6	7	8	9	10	Total
Number	788	722	656	656	656	656	656	656	656	525	6,627
Proportion	0.12	0.11	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.08	

New Zealand

The data from New Zealand differ from the rest of the data in a number of ways. The first is that children start school just after their fifth birthday and so arrive at school throughout the year. They do PIPS just after arrival and then one year later, rather than at the end of the academic year.

The number of schools in the sample from New Zealand is smaller than the datasets for either England or Scotland and national pupil level data on home background were not available. But school level information was available, in the form of the decile rankings for state schools participating in PIPS (New Zealand Education Counts 2013). Details of how the deciles, which are based on demographic information, are calculated are published (New Zealand Deciles 2013). Table 25 shows the distribution of the New Zealand schools represented in the PIPS dataset.

Table 25: Decile rankings for New Zealand schools participating in PIPS

Decile	1	2	3	4	5	6	7	8	9	10	Total
Number	2	1	1	1	2	3	1	1	4	6	22

All Deciles are covered by at least one school in the PIPS dataset. To gain a sample of students with an even distribution for New Zealand two schools were selected randomly for Deciles 6, 9 and 10. Deciles 2, 3, 4, 7 and 8 have only one PIPS school per Decile. The data for these Deciles are replicated to simulate two schools. Deciles 1 and 5 have two PIPS schools per Decile. This sample was used for further analyses.

Table 26: Number of pupils in each Decile in Primary 1 in PIPS sample New Zealand 2012 matched to two PIPS schools per Decile

Decile	1	2	3	4	5	6	7	8	9	10	Total
Number	64	70	32	38	140	88	88	138	88	101	847
Proportion	0.08	0.08	0.04	0.04	0.17	0.10	0.10	0.16	0.10	0.12	

PIPS pupil sample for Australia (one territory and one state)

PIPS data are collected in all Australian States and Territories but, for one territory and one state (A & B), data are available for all schools. This report only includes these complete datasets where there were 46 and 65 schools in A and B respectively. Because of an agreement within Australia, the Independent, state and catholic schools data cannot be separated and so, unlike England, Scotland and New Zealand the Australian data includes some independent schools. These data amount to 32% in A and 8% in B and all of the data were used in the analyses which follow.

Appendix 2: Reliability and Predictive Validity

Reliability

All of the data were initially analysed together using Rasch measurement in the Winsteps (version 3.74.0) package. The reading passages (which range from simple sentences to longer passages) were dichotomised into those which were scored very highly and those less highly. The Rasch model showed very strong psychometric properties. There were extremely reliable person measures (alpha=0.93) with good separation (around 4), extremely high reliability of the items (0.99) and excellent separation (around 30). Just a few of the items were not reached by any pupils and just a few pupils had such extreme scores that they could not be measured. The fit statistics for the items were good, with the Infit mean Square averaging 0.98 with a standard deviation of 0.1. This suggests most items fitted well but that a few either over or under fitted.

More details about the total score and the subscales can be found below.

Table 27: Reliability and separation based on data from England for this report without EAL children

Name	Person reliability from Rasch (Alpha)	Person Separation from Rasch
Total	0.93	4.00
Early mathematics	0.90	2.94
Early reading	0.84	2.32
Picture vocabulary	0.78	1.87
Phonological awareness	0.73	1.64

Additional data on reliability from the PIPS Technical Report (2001) is given in the table below. These figures add to an understanding of PIPS because they are based on reassessments whereas the table above relates to internal consistency

Table 28: Test/retest reliability abstracted from the PIPS Technical Report (2001)

Name	Test retest
Total (without phonological awareness)	0.98
Early mathematics	0.90
Early reading	0.97

N=29

Predictive validity[19]

As the years go by, more and more data is collected and this can be related to later assessments. In the tables below we can see the correlations between the Baseline score and the End of Reception assessment in PIPS based on the reading and the maths scores and correlations up to the PIPS reading and maths scores at Year 2, 4-5 terms after End of Reception, and also in relation to KS1 results. The results are based on the CD based version of PIPS. In general these findings give considerable confidence in the predictability of attainment; at least to the extent that anything can be expected to be predictable at this particular stage. Reports which give further details of the connections between the PIPS baseline assessments and later assessments are available from the CEM Centre.

Table 29: Test/retest reliability

	Correlation with PIPS total	Number of pupils	Date
End of reception maths	0.67	14460	94/95
End of reception reading	0.75	14460	94/95
Y2 maths (PIPS)	0.59	2512	97/00
Y2 reading (PIPS)	0.61	2516	97/00
KS1 reading task level	0.56	5975	97/00
KS1 reading comprehension level	0.55	6137	97/00
KS1 writing level	0.60	6964	97/00
KS1 spelling level	0.48	6790	97/00
KS1 maths level	0.59	6989	97/00
KS1 average level	0.66	7045	97/00

[19] This section is taken from the PIPS Technical Report 2001

Appendix 3: Differential Item Functioning

Differential item functioning (DIF) were therefore conducted for age, gender, EAL and deprivation. For each item, the size of the difference between each subgroup was inspected and the difference was tested for statistical significance. Because the sample sizes are so large in the majority of cases the former were taken as the most important and only if the size of the difference was greater than 0.5 logits was the p value inspected for t values which tested the difference between the average score and the score for a particular group. If that was less than 0.01 then it was concluded that there was DIF for that item.

Appendix 4: Pre-school

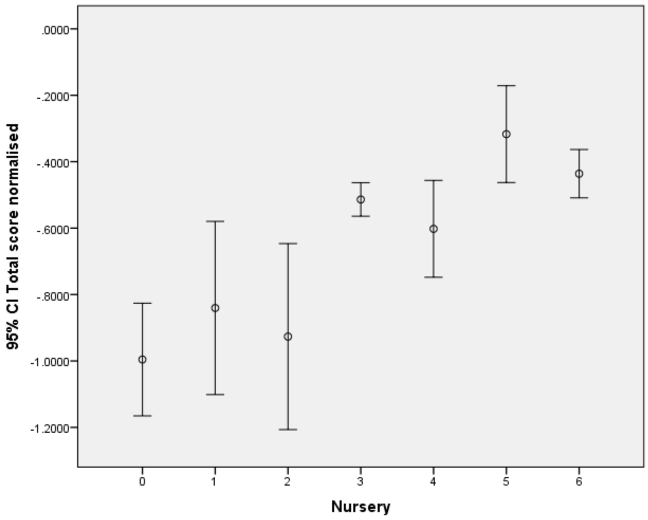
The numbers of terms at pre-school, as entered by the teacher into the contextual data section of the software are shown in Table 30 below. These items of data were provided on an optional basis and so limited samples are available.

Table 30: Numbers of children by numbers of terms at Playgroup or Nursery

Number of terms	Playgroup	Nursery
0	57	84
1	9	24
2	11	41
3	119	1,067
4	37	123
5	21	101
6	112	388

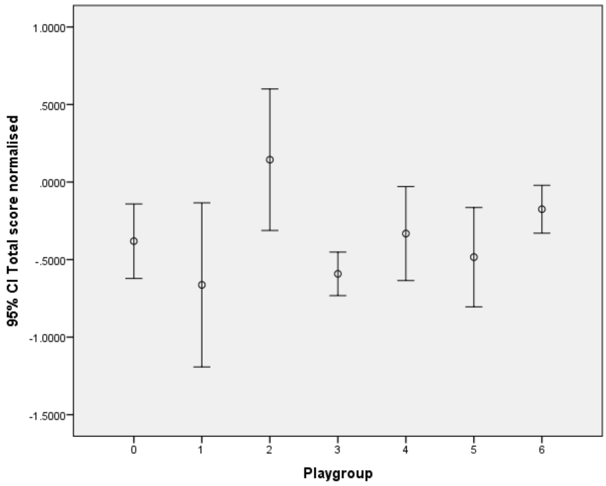
The graphs below show how these categories relate to normalised total score.

Figure 16: Number of terms at nursery related to the total score



The mean Total score in plotted on the Y axis with 95% Confidence Intervals

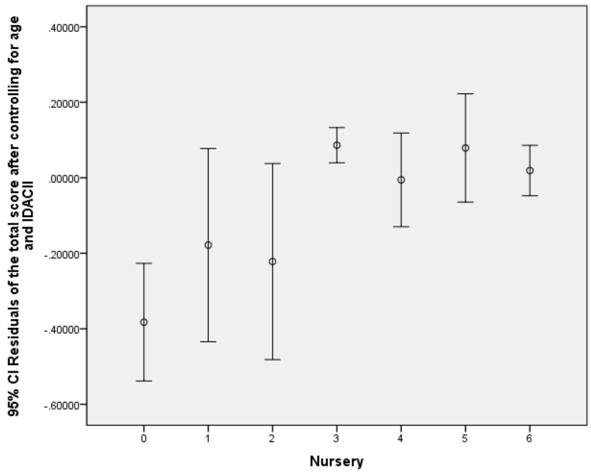
Figure 17: Number of terms at playgroup related to the total score



Visually, there appears to be little link between the standardised scores and attendance at play group (Figure 17) although when broken down by sub-measures there was a significant relationship for phonological awareness ($p<0.01$) which remains after controlling for age and deprivation. But, the link was not manifested as a clear trend and was not significant for early reading, early mathematics or picture vocabulary.

On the other hand, Figure 16 does suggest a trend related to nursery attandance and after controlling for age and deprivation a relationship between terms at nursery remains significant ($p<0.01$) and this held true for all sub-assessments. However, the trend appears not to be monotonic as the chart below shows.

Figure 18: Number of terms at nursery related to the total score after controlling for age and deprivation



Although based on limited data the chart suggests that one or two terms has little impact and that three or more terms are similarly positive. Further analysis does not suggest that that the link varies across deprivation groups.

Attendance at nursery was also link to PSED and was initially dichotomised into less than a year or more than a year on the basis that that was the cut off which mattered for the link to cognitive measures. But this did not produce a significant result ($p>0.01$) perhaps because the statistical test was underpowered as not all teachers chose to rate children on the PSED items and nursery attendance data were not available for all pupils. Further, a more detailed approach suggested that pupils with five or six terms at nursery were indeed scoring more highly on many of the items. Therefore the direct correlation with the total numbers of terms at nursery was used and this was converted to an effect size for the figure. For all items there was a positive link and it was significant for more than half of the items with modest effect size.

Table 31: Links between PSED and nursery attendance in effect sizes

Measure	Terms at nursery
Adjustment Comfortable	0.23**
Adjustment Independence	0.23**
Personal Confidence	0.19
Personal Concentration (self-directed)	0.28**
Personal Concentration (teacher-directed)	0.26**
Personal Actions	0.30**
Social Relationship (to peers)	0.17
Social Relationship (to adults)	0.22**
Social Rules	0.18
Social Cultural Awareness	0.18
Social Communication	0.24**

Based on 661 cases.
** $p<0.01$

Appendix 5: Dimensionality

An interesting question arises as to whether the data represent a single scale or if it should be broken up into more dimensions. Figure 19 below is designed to help answer this question and gives the results of a principal components analysis of the residuals from the Rasch model. They clearly suggest **one basic dimension dominating the data** but it is also clear from the Eigen value for the first contrast, which is 4.8, that a second dimension is present. There is also evidence of a third dimension and more.

Figure 19: Dimensionality of the total measure

```
INPUT: 168 ITEM  REPORTED: 6458 PERSON  WINSTEPS 3.74.0
--------------------------------------------------------------------------------------

     Table of STANDARDIZED REffect SizeIDUAL variance (in Eigenvalue units)
                                                 -- Empirical --    Modeled
Total raw variance in observations     =        292.5 100.0%         100.0%
  Raw variance explained by measures   =        127.5  43.6%          44.3%
    Raw variance explained by persons  =         45.2  15.4%          15.7%
    Raw Variance explained by items    =         82.3  28.1%          28.6%
  Raw unexplained variance (total)     =        165.0  56.4% 100.0%   55.7%
    Unexplned variance in 1st contrast =          4.8   1.6%   2.9%
    Unexplned variance in 2nd contrast =          3.3   1.1%   2.0%
    Unexplned variance in 3rd contrast =          2.9   1.0%   1.7%
    Unexplned variance in 4th contrast =          2.8   1.0%   1.7%
    Unexplned variance in 5th contrast =          2.5    .8%   1.5%

       STANDARDIZED REffect SizeIDUAL VARIANCE SCREE PLOT

        VARIANCE COMPONENT SCREE PLOT
       +--+--+--+--+--+--+--+--+--+--+--+
   100%+  T                             +
       |                                |
  V 63%+                                +
  A    |           U                    |
  R 40%+     M                          +
  I    |                                |
  A 25%+        I                       +
  N    |                                |
  C 16%+                                +
  E    |        P                       |
    10%+                                +
  L    |                                |
  O  6%+                                +
  G    |                                |
  |  4%+                                +
  S    |                                |
  C  3%+                                +
  A    |                                |
  L  2%+              1                 +
  E    |                                |
  D  1%+                 2              +
       |                    3  4  5     |
   0.5%+                                +
       +--+--+--+--+--+--+--+--+--+--+--+
          TV MV PV IV UV U1 U2 U3 U4 U5
         VARIANCE COMPONENTS
```

The first contrast grouped the identification of numbers and letters together. This possibly corresponds to children who have learnt letter and number identification in pre-school or home, but it could also be that some children are particularly attracted by symbols. This set of items is juxtaposed with another group of items from vocabulary, phonological awareness, Ideas about Maths and Ideas about Reading. Interestingly, the same pattern appears whether EAL children are included in the analysis or not.

The second dimension picks out three and two digit identification as well as the reading of some sentences.

The third dimension largely picks out reading passages as one group and maths items as another group.

There has been much debate in the literature about which scale to use when given statistics such as those derived above. The consensus seems to be that decisions must be influenced by the purpose of the measures and not solely based on quantitative indicators. In this case, we are initially interested in describing the general cognitive development of children when they start school, and in particular, how this relates to later literacy and numeracy. None of the groups from the first or second contrasts correspond well to the variables which the assessment was designed to measure although the third dimension does seem to split early reading and early mathematics.

The approach adopted here was to first set out a map of items and individuals and then explore the other scales, with a particular focus on the development of literacy and early mathematics.

Appendix 6: Measures for PSED

Adjustment – Comfortable

1. Upset on separation from carer at the start of the session. Not at ease during the day. Does not cope easily with transitions between activities or locations within the school setting.

2. Sometimes upset on separation from carer at the start of the session. Fairly settled during the day. Occasionally finds transitions a problem.

3. Rarely upset on separation from carer at the start of the session. Copes well with transitions between activities or locations within the school. Fairly settled during the day.

4. Never upset on separation from carer at the start of the session. Comfortable for most of the time during the session. Has no difficulty coping with transitions.

5. Never upset upon separation from carer at the start of the session. Very comfortable, never ill at ease during the session. Has no difficulty coping with transitions between activities or locations.

Adjustment – Independence

1. Dependent on adults or another child for guidance and support for much of the time. Generally needs help with clothing and personal activities (coat, toilet, etc.).

2. Some dependence on adults or other children. Needs help with some clothing and personal activities (coat, toilet, etc.).

3. Independence of others for most of the time but still needs occasional support. Can cope with some clothing and personal activities, but not all (e.g. can put on coat but is unable to fasten it).

4. Independence of others for most of the time but still needs occasional support. Copes well with most clothing and personal activities.

5. Independence. Seeks assistance only when special help is required. Can put on and fasten coat, go to the toilet, etc.

Personal – Confidence

1. Very hesitant. Does not join in group activities and rarely talks.
2. Fairly hesitant. Reluctant to participate in group activities or talk.
3. Will join in group activities or talks when prompted.
4. Quite confident. Keen to join in group activities or talk within the school setting.
5. Very confident. Keen to participate in group activities within the school.

Personal – Concentration (Self-directed activities)

1. Finds it extremely difficult to concentrate. Very rarely settles to one thing and very easily distracted.
2. Short concentration span. Finds it difficult to settle down to one thing. Easily distracted.
3. Able to settle to a task and concentrate for a sustained period. May be distracted.
4. Attends quite well. Able to maintain concentration and is not disturbed by mild distractions.
5. Can focus attention, even in the face of competing activities. Has been seen to concentrate for a long period (e.g. 15 minutes).

Personal – Concentration (Teacher-directed activities)

1. Finds it extremely difficult to concentrate. Very rarely settles to one thing and very easily distracted.
2. Short concentration span. Finds it difficult to settle down to one thing. Easily distracted.
3. Able to settle to a task and concentrate for a sustained period. May be distracted.
4. Attends quite well. Able to maintain concentration and is not disturbed by mild distractions.
5. Can focus attention, even in the face of competing activities. Has been seen to concentrate for a long period (e.g. 15 minutes).

Personal – Actions

1. Acts impulsively without any consideration for the well-being of themselves and others. Demonstrates inappropriate behaviour in all situations. Unable to cope with changes in routine.

2. Occasionally considers the well-being of themselves and others before acting, but still exhibits frequent impulsive behaviour. Will interact appropriately with others when prompted, sharing and taking turns.

3. On about half of all occasions, considers the well-being of themselves and others before acting. Sometimes interacts appropriately with others but still needs frequent prompts. Copes with changes in routine reasonably well but sometimes gets over-excited.

4. Frequently considers the well-being of themselves and others before acting. Usually interacts appropriately with others without being prompted to do so. Copes quite well with changes in routine.

5. Almost always considers the well-being of themselves and others before acting. Unless severely provoked, always interacts appropriately with others without being prompted to do so. Responds positively to changes in routine.

Social – Relationship to peers

1. Finds it difficult to communicate with other children and make friends. Seems to take no account of others and is frequently inconsiderate.

2. Often has difficulty communicating with other children and making friends.

3. Communicates quite easily with other children and able to form friendships. Takes notice of the feelings of others when they become very obvious.

4. Communicates quite easily with other children and readily forms friendships. Takes notice of the feelings of others.

5. Communicates very easily with other children and readily forms friendships. Is aware of others and responds to their needs. Sensitive.

Social – Relationship to adults

1. Finds it difficult to communicate with adults. Does not approach adults or speak to them. Inappropriate behaviour whilst interacting with adults.

2. Often has difficulty communicating with adults. Reluctant to approach adults or speak to them. Usually interacts appropriately with adults.

3. Communicates with adults but some difficulty. Will approach adults and speak to them. Rarely demonstrates inappropriate behaviour whilst interacting with adults.

4. Confident approaching adults when necessary. Relates well to adults, and with appropriate behaviour.

5. Confident approaching adults when necessary. Relates easily to adults, and with appropriate behaviour. Speaking to adults is natural and easily understood.

Social – Rules

1. Takes no notice of rules. Distracts others and interrupts activities.

2. Takes little notice of rules. Can distract others and interrupt activities.

3. Sometimes ignores rules. May distract others on occasion.

4. Usually obeys rules and rarely distracts others.

5. Always obeys rules and never distracts others

Social – Cultural awareness

1. Shows an awareness of the routines in their home environment. For example, through role-play in the ‘home corner’.

2. Demonstrates an understanding of being a member of a family/household by talking about relationships and experiences with parents/siblings and other relatives.

3. Is aware that they are a member of a wider community within their local neighbourhood and pre-school setting. Talks about experiences relating to those environments.

4. Recognises that the way of life of others may be different from their own. Takes pride in their own achievements.

5. Is aware of, and respects the way of life of others. Understands that their own way of life should be respected by other children and adults

Social – Communication

1. Communicates with others using single words, gestures and facial expressions.

2. Speaks using simple statements. Uses intonation to ask questions rather than grammatically correct language.

3. Begins to combine statements to present a coherent argument or explanation. Spoken sentences are generally a combination of ideas and not usually grammatically correct.

4. Asks simple questions. Spoken sentences are sometimes grammatically correct.

5. Speaks fluently and coherently. Speech is generally but not always grammatically correct. Listens attentively to the views of others and responds appropriately, taking turns in the conversation.

Department for Education

Reference: DFE-RR344

ISBN: 978-1-78105-354-6

Any enquiries regarding this publication should be sent to us at: Emily.KNOWLES@education.gsi.gov.uk or www.education.gov.uk/contactus

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