Network on Early Childhood Education and Care

iPIPS: AN INTERNATIONAL STUDY OF CHILDREN’S FIRST YEAR AT SCHOOL

Project for Discussion with the OECD Network on Early Childhood Education and Care

17-18 December 2012
Conference Centre Room CC10
OECD, Paris, France

This document is only available in PDF format.

This paper was prepared by Professor Peter Tymms, Dr. Christine Merrell and Professor David Hawker (Durham University, United Kingdom) for the 12th meeting of the OECD Network on Early Childhood Education and Care held from 17-18 December 2012 at OECD headquarters in Paris, France.

Miho Taguma, Senior Policy Analyst; Tel: +33 (0) 1 45 24 92 65; Email: miho.taguma@oecd.org

JT03331577

Complete document available on OLIS in its original format

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.
Why do we need an international study of children’s first year at school?

It is universally recognised that children’s early development, and their progress during the first year of school, are crucial for their later success.

Policy makers everywhere are interested in knowing how well their children are doing at this key stage in their life, not just in their own country, but also in relation to others.

Questions such as:

- How does the development of children in my country compare with other countries at the start of school?
- How do children compare after a year in school?
- What do these differences say about the effectiveness of Early Years policies in my country by comparison with other countries in similar circumstances?
- What do these differences say about how well children learn in their first year of school in my country?
- To what extent are differences in later international surveys (e.g. PISA, PIRLS and TIMSS) explained by differences in the Early Years?
- How strong are the links between development and socio-economic status in my country and how does that compare with similar jurisdictions?
- To what extent can differences in children’s starting points be attributed to differences in pre-school provision and home environment over and above socio-economic status within and across countries?

The International Performance Indicators in Primary Schools project (iPIPS) offers schools and policy makers the chance to help answer these questions. For the first time, it will provide accurate benchmarks of the attainment of children starting school on an international basis. Over time, it will generate stronger and stronger evidence to enable national and regional authorities to compare their Early Childhood Education and Development approaches and innovations against their earlier benchmarks and those of others, and thus directly contribute to the improvement of early education.

Why is this study sensible and viable when suggestions for similar studies have been rejected?

- The study is based on a well-established monitoring system (PIPS) which was originally developed 18 years ago in the UK, and has been successfully used for over 15 years in a number of countries.
- The PIPS team based at the Centre for Evaluation and Monitoring (CEM) at Durham University, UK, has extensive specialist expertise in both Early Childhood Development and survey based assessments.
- PIPS has proven success in collecting valid data on children starting school in different cultural contexts and different languages, (including Abu Dhabi, Australia, the UK, Germany, New Zealand, the Netherlands, South Africa) and is therefore suitable for adaptation for a sample based comparative international study.
- Reliable data concerning children’s cognitive development is collected in a 20 minute computer adaptive one-to-one assessment.
- Teachers’ ratings of children’s personal and social development are a key feature, and the questionnaire instrument to collect this information has been developed and tested over more than a decade.
The survey also includes the collection of basic data on children’s physical development, and contextual data, to allow efficient analysis of how all these factors interrelate.

The survey includes a follow up assessment at the end of children’s first year in school, enabling measurements of relative progress to be made.

Technical solutions have been found to the challenge of providing valid assessments and comparisons of children’s development when they start school, while allowing for the range of different school starting ages in different countries.

The experiences of PISA, TIMSS and PIRLS provide additional models to help inform the development of iPIPS.

League tables will not be produced. Rather each country will get a report which contextualises their results against comparable data. Robust measures will be taken, in the design of the assessments and the presentation of the results, to avoid the survey becoming a ‘high stakes’ procedure for individual schools or teachers, or for the assessments to exert undue influence over the nature and content of pre-school learning programmes.

**Project for Discussion with the OECD Network**

This proposal is directly relevant to the ECEC network’s proposed policy output on early learning and development, as set out in the paper presented to the meeting held on 11-12 June. Specifically it could provide some of the high quality information envisaged under strand 5 of the program: ‘Data, research and monitoring’. The first question posed in the paper under this strand is ‘Which data would still need quality improvement, of the existing international data?’ We believe that an important gap in the current international data is the lack of any robust comparative baseline assessment of children’s attainment as they start school. iPIPS is specifically designed to fill that gap.

As outlined above, we believe that the PIPS assessment which forms the basis of iPIPS is well proven in terms of its reliability and predictive validity, making it robust enough in principle to be used in a cross national context. We also believe that we have found answers to the vexed question of measuring the attainment and progress of children in systems with very different policies concerning school starting age and early years provision. We believe therefore that it will be fit for the purpose of providing the comparative data which are currently lacking.

The proposal comes from the Centre for Evaluation and Monitoring (CEM) at Durham University, UK. It is supported by a number of international partner organisations including ETS and ACER (Australian Council of Educational Research).

Several jurisdictions have already used the proposed assessments for many years although, with some exceptions, their use has not normally involved representative samples of children. The computer-adaptive assessment has been administered in participating schools, usually by class teachers. It has included measures of early reading, early mathematics, measures of executive function, personal and social development, behaviour and attitudes to school and learning. Once administered, the data are uploaded via a secure website to a central processing unit at CEM and standardised feedback returned to schools.

The development and progress of children starting school in different jurisdictions have been investigated, and findings published in academic literature. Building upon this work, CEM and its partner organisations now propose to work together to develop iPIPS into a large scale international survey.

This work will take place in three overlapping phases, as explained below.
Phase 1: 2012-13 (Preparatory phase)

Building upon past work, this phase is currently underway and involves:

- Scoping of the study with international partner organisations, including exploration of key methodological and policy issues.
- Recruitment of up to six pilot jurisdictions to provide trial sites for testing the materials, technical standards and procedures. At the present time, the jurisdictions that we are talking to are Abu Dhabi, Australia, England, Germany, Hong Kong, New Zealand, Scotland, South Africa and the USA, from which we anticipate six being in a position to participate in Phase 1.
- Joint work with representatives of these jurisdictions to develop the materials and procedures.
- Further analysis of existing PIPS datasets to add to the existing research literature with material specifically relevant to iPIPS.
- Benchmarking studies with other early years’ assessments to establish points of comparability and difference between iPIPS and other instruments.
- Technical studies and consultations with OECD Network members and others.
- Establishing an international advisory board.

Phase 2: 2013-15 (Pilot Phase with Six Jurisdictions)

- Administration (in 2013-14) of baseline surveys in six jurisdictions. The first assessment will take place in September/October 2013 in jurisdictions in the Northern hemisphere and in January/February 2014 in the Southern hemisphere. Children in New Zealand start school throughout the year when they turn five years and so assessments will be conducted during the full year.
- Production of trial jurisdiction reports.
- Development of extension material.
- Evaluation of trials and refinement of procedures in the light of the trials and consultation.
- Preparation and administration (in 2014-15) of one-year-on follow up surveys in trial sites; general invitation issued for jurisdictions to take part in first ‘live’ round.

Phase 3: 2015-17 (Scale-up)

- Scale-up of baseline surveys, starting in September 2015 (Northern hemisphere) or January 2016 (Southern hemisphere) with follow-up at the end of the year (July/December 2016).
- Feedback to schools and parents within three months of data collection against the background of each jurisdiction separately. This is both for the start of year and the follow up.
- Start-of-school individual jurisdiction reports and international summary report (2016).
Costs of participation

The cost of participation in Phase 2 will be calculated on the basis of the work required to adapt, and translate the assessment and surveys, software development, sampling framework, data analysis and reporting. Local costs of administration and data collection would be additional.

iPIPS will be derived from the existing PIPS assessments, therefore the development costs for each jurisdiction taking part will be relatively modest. Essentially they consist of carrying out translations of all the materials where necessary, adaptations to some parts of the assessments (principally the reading literacy and vocabulary sub-scales), trialing costs, and sample design. In terms of the administration of the assessments, there will be costs involved in drawing the sample, training the administrators, and collecting and analysing the data. Because the assessments are computer-based, these aspects of the exercise are relatively inexpensive.

It is not possible to give a precise figure in advance for the costs of administering the assessments in any given jurisdiction. This will be for each participating jurisdiction to determine according to prevailing local factors. We do, however, expect that, in most contexts, iPIPS is likely to prove substantially less expensive than other international benchmark surveys. Further details can be found in Appendix 1.

Management and governance of iPIPS

Phases 1 and 2 of iPIPS will be managed by CEM at Durham University, in collaboration with the international partner organisations and the six pilot jurisdictions. An international advisory board will be established, comprising these partners, together with representatives of international organisations (including prospectively OECD, the World Bank and IEA) and leading independent experts.

Professor Peter Tymms is the iPIPS Director. He will chair the international advisory board and co-ordinate a central management group, which will be based at CEM.

For Phase 3, a consortium will be set up with expert sub-groups to deal with:

- Sampling methodology, standards and procedures.
- The content and design of the data collection instruments (assessment materials and questionnaires).
- Translation and adaptation of the materials.
- Liaison with participating jurisdictions, including administrative procedures.
- Analysis of data.
- Reports.

The management and governance of iPIPS will be reviewed during Phase 2, with a view to putting longer term arrangements in place for Phase 3 and thereafter.

Discussion of key questions

1. Why take part in an international baseline assessment survey?

There are a number of important policy questions concerning early education for which there is currently no comparable international data, despite the immense interest in all countries in the issues. There is no equivalent system to the OECD’s PISA assessments, or the TIMSS and PIRLS assessments carried out by IEA, for children starting school. iPIPS is intended to fill that gap. In terms of direct relevance to policy makers, assessments conducted at both the start of school and at the end of the first year will help to inform four areas of policy:
• Pre-school policy. The children’s assessment will be accompanied by a questionnaire survey of parents/carers to capture information on the kind of pre-school experience children have had, and for how long, whether formal pre-school, or informal childcare, full time, part time, or none at all. Questions will also be asked about interactions at home and key features of children’s development. The focussed survey will be carried out, where possible, face-to-face in the school setting. Together, this information will help to ascertain the impact of different forms of pre-school provision and experience on children’s development, in a comparative international context, so that national and regional authorities can decide how (and how much) to invest in their pre-school provision, and what types of provision are most effective internationally.

• School starting age policy. Different administrations have different policies on the age at which children start formal schooling, and at what point their attendance becomes full time. Although much has been written about the age of starting school it has not been possible to draw conclusive results. By comparing the results of the ‘on entry’ with the ‘one year on’ assessments, against similar data from administrations with different starting age policies, it will be possible to make more informed judgements about the effectiveness of different policies.

• Curriculum and pedagogical policies. School providers will be asked about their approaches to education during the first year at school. The results of this survey when linked to other data which are being collected will highlight, at the jurisdiction and global level, the relative success, in terms of the progress which different groups of children make, of different policies concerning the curriculum taught, and the pedagogical process. At a more detailed level, the survey results can also be used to identify schools or school districts, or types of approach within a system, which appear to be working more effectively than others, and thus help support school improvement.

• As a baseline for later surveys. Currently, the extent to which countries’ differential performance on international surveys such as PIRLS, TIMSS and PISA is due to differences in children’s starting points is not known. This survey would begin to provide context to differences seen. It is clearly important for all systems to have this information in order to gauge more accurately whether pre-school policies need to change in order to improve performance, or whether problems arise at a later stage, and therefore need to be addressed through school age policies. However, because there is a lack of suitable baseline data at present, the international discourse is currently characterised by assumptions (albeit potentially informed by, largely US based, intervention studies) about the power of certain types of early education in building a foundation for later success.

2. Who and what is being assessed, and how?

This is an assessment of children starting school that reports on their level of development in early reading and mathematics, their vocabulary development, their general personal and social development, and their executive function including short term memory. Some parts are administered through a series of short, engaging, computer based exercises, using adaptive testing techniques which are designed to focus on what children know and can do, and stop at the point at which children cannot do the questions. Some parts are assessed using observational rating scales. It is a highly efficient assessment, yielding the maximum useful information in the shortest time. Computer-delivered parent/carer and teacher surveys will collect background contextual data.

For the international survey, a sample of children will be assessed in order to yield aggregate results which are representative of the population of children starting school. Immediately following the initial sampling procedure, the assessment can also be used with all the children in a particular school or district, in order to yield more comprehensive local data which can be used to inform curriculum and teaching strategies, and individual data which will enable schools to address the learning needs of each individual child. The assessments have been designed to serve both purposes.
The assessment is designed to be re-administered at the end of children’s first year in school, in order to get a measure of progress at a cohort and individual level. The software tailors this second assessment to each pupil on the basis of their score at the start of the year to avoid repetition of easy material, thus ensuring a motivating and challenging experience for children and an efficient method of collecting data. This second assessment point is of key importance since it will allow policy makers to evaluate the progress their schools have made against international benchmarks. Together, data from the two time-points provide a robust baseline against which later assessments such as PIRLS, TIMSS and PISA can be better interpreted. Different national and regional administrations have different policies on the age at which children start school. Because of its adaptive nature, the assessment is suitable for use with young children aged from 4 to 7 years, of all abilities and with or without pre-school experience. These policy differences will be taken into account when the results are reported.

3. Why stage not age?

International surveys of achievement often concentrate on a particular age, but it is well known that the number of years of schooling is important to children’s development. In setting up a study of children at the start of their school careers, a key decision is whether the focus should be on age or stage. There are pros and cons to both these approaches.

If we were to assess representative samples of children aged 4 and then follow up them up one year later in a set of countries, this would provide valuable data for looking backwards in general terms at policies in the early years. It would also provide useful comparative data on the progress made by children between the ages of 4 and 5, independent of any particular regional or national policies concerning school starting ages. However, in most countries, pre-school provision is characterised by a variety of different institutional arrangements, which means that it would be difficult to identify the effectiveness of any particular type using a population based survey unless very large samples were envisioned. At the one-year-on follow up, some children will be in institutions including schools but some would still not be; the results would therefore confound maturation with educational provision and not necessarily reflect the major impact on children’s progress of their first year in school. The conclusions drawn from the findings would be limited.

It is evident from a number of studies (see references) that children’s academic progress is closely associated to schooling. This goes beyond traditional school effectiveness studies which concentrate on the differences in children’s progress between schools. Further, the number of years in schooling has been shown to be more important than age in determining reading levels and children’s development against other cognitive measures. Alexander and Martin (2004) write:

“The effect of one grade was over twice the effect of 1 year of age for most verbal cognitive ability subtests in Grades 4–6. A higher ratio might be expected for more curriculum-related areas such as reading tests.”

It is also clear that the teacher is a key factor in the progress made by children. This is particularly important in the first year of school, when children tend make more progress in relative terms than in any other year of their education.

An alternative approach to collecting information from all children at a particular age is to focus on children starting school, in order to capture the ‘school effect’ in children’s learning. Children in different countries might be as young as 4 or as old as 7 when they start formal schooling, although in most countries they start between ages 5 and 6. We do, however, have to define ‘formal schooling’. For the purposes of this study, we propose to define it as the point at which all or nearly all children are attending an educational institution on a daily basis where they will learn to read (even when this is not the official statutory school starting age). This definition thus caters both for systems in which parents overwhelmingly opt to send their children to school before the statutory
school starting age, such as the UK, and for those, such as Russia or Scandinavian countries, where the school starting age is much later, but where nearly all children attend a pre-school institution from the age of 5 or 6.

Nevertheless, even with this definition, children will be at a range of different ages across different jurisdictions when they start ‘formal schooling’. A study which focusses on children in this way therefore has the immediate challenge that children’s vocabulary, or any other measure that we care to pick on, will, on average, increase by age. We have, however, published data showing that it is quite possible to control for age and still be able to measure attainment on a meaningful basis. There are more or less linear relationships in children’s attainment in early maths, reading and vocabulary development for this age range across the sample of countries available to us. This allows us to make valid comparisons between children of different ages.

The chart below shows the early mathematics scores of children aged 4 to 7 starting school in England, Scotland, New Zealand and Australia. The y-axis shows the age rising on 3 month intervals from just 4 to nearly 7. There is a clear trend between age and development.

Further, since the institution and the teacher are of fundamental importance, it makes considerable sense to look at how school and teacher effects vary within and across countries, by carrying out the initial assessment at the start of children’s first year in school, and the follow up assessment at the end of that year.

In practical terms, when it comes to following the children up, it is much easier logistically to assess within one school and then one year later to pick up those same children in that school, with special arrangements to follow up the small proportion that have moved institutions during the year. Further, a school based assessment makes it much easier to collect data about the schools and their approaches as part of the survey. There is an additional key reason for collecting data from schools rather than homes: some of the most important data about children’s social, emotional and personal development needs ratings that come from observation over protracted periods. Collecting such data from parents is really problematic as we would be asking them to rate their own children, often in the absence of comparative information. Teachers have a much broader experience of similarly aged
children and, by use of videos, we have a straightforward way of standardising their assessments across and within cultures.

Finally some of the most important research questions require the analysis of high quality longitudinal data. For example, vexing questions about the optimum age for children to start formal schooling will, for the first time, be addressed with serious detailed data.

4. What comparative data will it yield?
The direct assessment of children will yield data on the following domains and sub-domains:

- Handwriting – the child is asked to write his/her own name.
- Vocabulary – the child is asked to identify objects embedded within a series of pictures.
- Ideas about reading – assesses concepts about print.
- Phonological awareness – rhymes and repeats.
- Letter identification – a fixed order of mixed upper and lower case letters.
- Word recognition and reading (words, sentences and comprehension).
- Ideas about mathematics – assessment of understanding of mathematical concepts.
- Counting and ability to use numbers.
- Sums – addition and subtraction problems presented without symbols.
- Shape identification.
- Digit identification.
- Mathematics problems – including sums with symbols.
- Short term memory

Measures of height, weight and other physical attributes (e.g. hearing loss) will be incorporated into the assessments for Phase 3.

Accompanying questionnaires, to be completed by parents and teachers, will also yield a range of valuable background information for each child, which will be matched against the outcomes data gained from their assessment. These questionnaires will collect contextual information, for example the pre-school home learning environment, use of local services, access to and uptake of local and national pre-school initiatives, uptake of pre-school provision, languages spoken in the home, number of siblings, parents’ education and socio-economic status.

Data from the parents’ and teachers’ questionnaires can be analysed in a number of different ways, to draw out detailed findings, such as the relative importance of various socio-economic factors on children’s development, and how this varies from country to country. A number of research studies have looked at these factors over the years, and it is in many ways a well-trodden path. For example, the Effective Provision of Pre-school Education (EPPE) investigated the impact of pre-school provision and home learning environment on children’s later outcomes. However, it followed a single cohort in England and therefore cannot provide information about the impact of policy changes over time and is of limited value internationally. However, to date there has been no systematic international benchmark data to confirm or challenge existing research. Collecting data from parents on uptake and satisfaction with pre-school provision will enable policy makers to judge the impact of policies from a different perspective to the system-level indicators of facilities, spending, quality, etc. available from resources such as Education at a Glance, Starting Strong 3 and Eurydice. It is important to note that we appreciate the extensive work already done in this area by the OECD and we would welcome engagement in the development of the surveys for use in this project.
In total, iPIPS will provide such evidence, and thus provide policy makers with a more powerful basis for making policy decisions than previous research studies have been able to do.

5. What is the proposed sample size per jurisdiction?

We estimate that a sample size of 3,000 pupils per jurisdiction will be sufficient for most purposes. However the precise size of the sample will depend upon the characteristics of the population to be surveyed and the number of separate sub-group analyses which are to be carried out. We intend to finalise the sample size in discussion with each jurisdiction, based on what their precise policy interests are.

6. How robust will the data be?

In technical terms, the data will be very robust. The PIPS assessment instruments have been developed and refined over 18 years, both in the UK and elsewhere, and the individual assessments of children have been shown to have high reliability, with inter-rater correlations of 0.98 and high predictive validity, with correlations up to 0.68 with later assessments of children at age 11. It has been translated into 10 languages, and has been proven to work in countries where traditions of pre-schooling, and modes of learning, are very different from the UK.

The sampling processes and other procedural aspects of the assessments will conform to internationally recognised technical standards. In addition, an independent sampling referee will oversee the sampling procedures used in each participating administration, as this will be a crucial factor in securing the validity of the survey outcomes.

The assessments themselves will be thoroughly trialed in each administration where they are taking place for the first time, to ensure that they are culturally and linguistically appropriate, while retaining a high level of international comparability.

7. How soon will we get useful analyses?

Information about individual jurisdictions will be made available to participating administrations very quickly after the administration of the assessments. The analytical frame will be set up in advance, in collaboration with the participating jurisdiction, so that, once the results are in the database and validated, some analyses can be produced almost immediately. Because the assessments will be computer administered, the data will be entered simultaneously, and will be available for analysis as soon as the necessary validation checks have been carried out.

The speed at which international comparative data is released will depend on how fast the assessments are administered in each participating jurisdiction. In Northern hemisphere countries, the assessments (including the school and parent/carer surveys) will be completed by December, and in the Southern hemisphere by April, so the expectation would be that international benchmark data would be available to the international advisory committee one month later, i.e. January in the Northern hemisphere, May in the Southern hemisphere. Subsequent decisions concerning the publication of results analyses, when and how, will be made by the international advisory committee although indicative dates are given above. Decisions concerning the policy use of the results will of course rest with the individual participating jurisdictions.

The ‘one year on’ assessments will take place towards the end of the children’s first year in school and a similar timescale will operate then.

8. Who else has used this survey and what policy decisions has it led to?

Since this is a new international survey, it has not yet been widely used in this form by policy makers. However, the existing PIPS system has been widely used in the UK and in other countries for school improvement purposes. Several hundred thousand children have been assessed in thousands of schools. Through providing almost instant diagnostic feedback on individual children, it is used by teachers and school principals to refine their teaching
strategies to target areas where children need additional help. At the school level, the PIPS database already provides comparative data which helps administrators to identify those schools where children are less advanced in their learning than elsewhere, and to take remedial action at a whole school level, for example by providing additional expert support, changing teaching methods, addressing performance management issues, etc. The second administration of the assessments one year on will be a particularly powerful tool for doing this.

In two countries, however, the data provided by PIPS has been explicitly used for policy purposes.

In Scotland PIPS has been used extensively over a number of years to assess children’s abilities on entering school. A report funded by the Scottish Government in 2005 (Children Starting School in Scotland, SEED 2005) analysed the results against a number of background factors, including pre-school experience, age of starting school, sex, and home background. It also used comparative PIPS data from England, Western Australia and New Zealand too as benchmarks for comparing the Scottish results. The purpose of the report was to help the Scottish Government shape its policy on pre-school and early years education, in particular decisions about the best age for children to start school.

PIPS is used at state level by the Abu Dhabi Education Council (ADEC) who are committed to a transformational reform agenda with the aim of elevating student performance to international standards. One of their key areas of focus has been the standards of education in Kindergarten, Grade 1 and Grade 2. CEM has worked in partnership with ADEC to translate PIPS into Arabic for use with these students. The assessment is being used across the Emirate to monitor the effectiveness of their reforms and the first report has shown a marked improvement in areas where their interventions have been targeted. See: http://www.thenational.ae/news/uae-news/education/maths-and-literacy-to-be-tracked and http://www.edarabia.com/26751/maths-and-literacy-at-abu-dhabi-primary-schools-to-be-tracked/ for more details.

The intention of iPIPS is that, through adapting the assessments for use in a range of different countries, the system level information on children starting school which has been used effectively in Scotland and Abu Dhabi will become available to a wider range of countries, and thus help to inform the relative effectiveness of different early years’ policies.

9. This would be the first international attainment study with a true longitudinal component, re-testing the same children. What new technical challenges does this represent?

As noted PIPS has already been used in several countries over many years to assess children’s development at the start of school and to measure their progress in the first year. This includes English-speaking countries such as Australia, England, New Zealand and Scotland, and others such as the Netherlands and Germany. The CEM Centre has established a working model for the longitudinal assessment of children and has published research on comparisons in some developmental areas. The computer-adaptive assessment can be adapted to different cultures and languages relatively easily and there is a secure on-line facility for the upload of assessment data and download of feedback. Several thousand schools have used this facility without problems.

The iPIPS team is working with experts at the Australian Council for Educational Research (ACER) and ETS on the technical aspects of sampling.

10. Will the results of iPIPS be presented as international league tables?

League tables of international performance have been used to make direct comparisons of quality and outcomes between one national system and another for PISA, TIMSS and PIRLS assessments, where children have been in school for a considerable time, and their performance on key areas of the curriculum can be compared more or less on a like-for-like basis.
The same cannot validly be presented for the early years, where different national policies result in children starting school at different ages, having experienced very different types of pre-school provision (or none) and, crucially where it is not yet clear what the best age of starting school is or indeed if that varies according to language and script. Further, the intended outcomes in different systems are often very different. Some systems concentrate heavily on developing children’s early literacy and numeracy, while others focus on developing their personal and social skills, deliberately leaving formal instruction until later. This is contested territory, and as yet there is no consensus internationally as to what kinds of approach are best. There is a danger that presenting results in the form of simple tables of comparative results would have a narrowing effect on the international debate concerning effective early education, precluding other factors, such as building social capital in families, which are less amenable to measurement.

For these reasons we do not propose to present the results of the iPIPS study in the form of rank orderings of jurisdictions in, for example, children’s reading or mathematics skills. Rather, we would want to present a more rounded picture of children’s wider development, as this relates to the early years policies being pursued by the jurisdiction in question, so that jurisdictions themselves can decide on the most appropriate comparisons to make on a like for like basis. We do, however, expect to report on how children in each jurisdiction progress in comparison to similar children elsewhere. Importantly, the iPIPS assessment will yield a ‘double baseline’ (on entry assessment and follow up assessment) which will allow a measurement of value added to later international assessments. Using this baseline approach, future rounds of later assessments will be able to incorporate an additional type of benchmark — based on value added progress from the start of school - into their outcome reports. The combination of assessments will give jurisdictions a powerful set of indicators with which to judge the comparative effectiveness of their policies, (i) for pre-school provision, (ii) for the first year in school, and (iii) for children’s later academic progress.

11. How to avoid teaching to the test and other unintended consequences?

Despite good intentions, national and international assessments can easily become ‘high stakes’ for jurisdictions, and this can distort the measures used, either by encouraging teachers to ‘teach to the test’, or by exerting an undue influence over the content of the curriculum. It is of course axiomatic that any valid assessment should largely comprise items which reflect what children would typically be expected to be able to do at any particular age; however, no assessment is ever designed to do more than sample these learning domains, and there are many aspects of learning which cannot easily be assessed, even via a teacher questionnaire. The iPIPS assessments do not set out to cover the whole of children’s development; they are designed to provide accurate measurements against a small number of domains (around 30) which have proven over the years to have strong predictive value in terms of children’s later attainment.

The assessments are, therefore, designed to provide significant information about the skills and disposition which children acquire, but they are not definitive in terms of the curriculum which should be taught. This is an important distinction, which needs to be understood by both policy makers and teachers.

In order to avoid the assessments becoming ‘high stakes’ for schools or teachers (for example through the use of ‘league tables’ or performance management procedures), the reports will not include any data identifiable to an individual school or teacher. This should minimise the risk of schools (or pre-schools) ‘teaching to the test’ in order to obtain better results, either in the initial baseline assessment or in the follow up assessment. The iPIPS team will work with individual jurisdictions to ensure that this position is safeguarded. It is possible that additional measures may be included in the test (such as additional control items) which would be able to act as a mechanism for preserving the integrity of the survey.

At school level, therefore, there will be an ‘expectation of honesty’ based on the fact that schools and teachers are not being held to account. In addition, at least part of the data will be collected by independent researchers. The inclusion of new and unknown items in the follow up survey would also allow schools to be dropped from the analysis where there was evidence of inappropriate administration procedures.
12. The proposal assumes use of the existing PIPS instruments, simply translating them into other languages. English as spoken and written in England has very specific characteristics. Do you have a plan to minimise measurement error by ensuring that the test items transfer from one linguistic and cultural context to another?

We already have translations into a range of other languages (including German, Dutch, Slovenian, Arabic) which have been used and evaluated. We have equated these different versions of PIPS using Rasch measurement. Not surprisingly, the mathematics assessments show the most stability across cultures, but almost all the scales work pretty well. Inevitably, the reading and vocabulary scales present the most challenges although the other scales help us to measure the variability here, and give us the information we need to adapt these components so that they also generate comparable performances. So in terms of test design, we are confident that we have a viable way forward whilst acknowledging that equating vocabulary and early reading across all cultures and scripts may not be possible. The aim will always be to compare like with like.

13. How can you hope to get comparative data on ratings in different cultures with different behaviours and understandings?

This issue relates specifically to the assessment of children’s personal, social and emotional development, which is carried out by means of teacher ratings rather than an objective assessment. There may be considerable cultural differences in expected norms of behaviour between one country and another. For example what is perceived to be hyperactive behaviour in Japan may be seen differently in the US. In order to be able to carry out valid international comparisons, we need to ensure that we can benchmark teachers’ judgements across cultures whilst establishing the reliability and the relative severity/leniency of each rater. This will be done using video snapshots of children in different countries. We will use examples of children’s behaviour and instruments such as the American Psychiatric Association’s DSM IV for the teachers’ assessments of children’s development. All teachers involved in the project will be shown around six short clips from different cultural settings and their own, and asked to rate them on the iPIPS scales. They will also rate their sample of students. Using the Facets model in Rasch we will be able to establish the leniency and severity of the raters as well as their consistency (reliability). The exercise will also establish the cultural norms (leniency/severity) and whether indeed the construct can be reliably assessed within that cultural context. This will allow us to estimate the degree of error and bias within a system and establish reliable and valid comparisons.

14. What is the likely time burden on schools?

Each child completes a computer-adaptive assessment of early reading and early maths development, and this takes around 20 minutes per child. This can be done reliably by trained researchers and need not be a time burden for schools apart from organising time with the children. Additional teacher assessments of behaviour are likely to take five to ten minutes per child. Parental questionnaires about contextual factors would be administered by the project researchers rather than by school staff.

Bearing in mind that iPIPS is intended to be a sample assessment, we do not anticipate that the burden on any individual school will be very great. Large numbers of schools already buy into PIPS because of its value as a diagnostic instrument, so we would hope that this would continue to be a valuable optional extra for schools selected to take part in iPIPS.
15. Can you let us know more about the computer-based assessment and why it is the most appropriate mode of assessment for this age group?

More information can be found in the enclosed technical report for the English version of PIPS for information about the developmental areas assessed, the reliability and validity of the assessment. The PIPS Baseline Assessment has been used by schools since 1994 and has been found to have exceptional psychometric properties. In addition, schools have reported the experience to be a positive one for children and teachers alike.

An important extra consideration is that the decision making in response to the children’s reactions is controlled by the computer, which means that there are no errors in handling the data. This of course adds further to the reliability of the exercise.

The current PIPS assessments are administered on a PC platform. For Phase 3 and subsequent roll out, the software is being re-written for a range of different platforms, including tablets. It will be possible to administer the assessments either online or offline, with the data being submitted later. In this way we expect to be in a position to make the assessments available universally throughout the world.

References


Appendix 1: Further Information of Costing Model

Phase 2
- Versions of the iPIPS assessment software are created for use in six countries.
- Each participating country employs the assessment on a sample of 3,000 pupils across 50 schools, at the start and end of the school year.
- Development work and data processing to be the responsibility of CEM, whilst working with jurisdictions.
- The assessments will run during 2013/14 and 2014/15.
- CEM will prepare reports for jurisdictions in collaboration with each jurisdiction.
- CEM will generate feedback for schools.

Development Costs
These are the costs incurred by CEM to develop a customised assessment package for use in each jurisdiction. It is assumed that certain aspects of the development (e.g. trialing, translation in conjunction with CEM, providing sound files) will be carried out by each jurisdiction at its own expense.

The elements which need funding include instrument refinement, adaptation and trialing of assessment materials, technical quality management, analysis of psychometric functioning, development and agreement on sampling methods, frames and procedures, consultation and liaison, project management, establishment of steering board and stakeholder consultation forum, development of policy briefs, training and implementation, design of training package, administration of survey, data management, post survey work, review with participants, technical evaluation, evaluation with policy makers and report with recommendations to inform the next round.

Overheads
Costs incurred by CEM in overseeing and managing the assessments over Phase 2 across jurisdictions. It is proposed that these costs are split equally among the jurisdictions.

Running costs
Assuming sample sizes are the same, this element will again be split equally among the jurisdictions.

Phase 3
Mature versions of the iPIPS assessment run every three years in participating countries.

Further development of the assessment and system, as described for Phase 2, will be required for those jurisdictions who participated in Phase 2 in order to reach a stable version of iPIPS for each jurisdiction. Further development as described for Phase 2 will be required for new jurisdictions.

Possible Funding Model
Looking to a future in which iPIPS is a well-established system, it is proposed that a funding model is implemented by which wealthier nations (as measured by nominal GDP) pay proportionally more according to a formula. This could include a common fixed fee paid by all participants, with an adjusted formula being applied to the remaining funding. The amount of the fixed component needs to be decided upon.