



CALL FOR TENDERS

INTERNATIONAL EARLY LEARNING STUDY

100001420

The deadline date for the receipt of Tenders is Monday February 15, 2016 at 5.00 pm (Paris time)

INSTRUCTIONS TO TENDERERS

The OECD brings together the governments of [countries committed to democracy and the market economy](#) from around the world to:

- Support sustainable economic growth
- Boost employment
- Raise living standards
- Maintain financial stability
- Assist other countries' economic development
- Contribute to growth in world trade

The OECD also shares expertise and exchanges views with more than **100 other countries and economies**, from [Brazil](#), [China](#), and [India](#) to the least developed countries in Africa.

Fast facts

Established: 1961

Location: Paris, France

Membership:

34 countries

Budget: EUR 357 million (2014)

Secretariat staff: 2 500

Secretary-General:

[Angel Gurría](#)

Publications:

250 new titles/year

Official languages:

English/French

Monitoring, analysing and forecasting

For over 50 years, the OECD has provided statistical, economic and social data comparable with the most important and most reliable in the world. In addition to its collection of data, the OECD monitors trends, analysis, and forecasts economic developments. The Organisation studies changes and developments in trade, environment, agriculture, technology, taxation and more.

The Organisation provides a setting where governments can compare their experiences in developing public policies, seek answers to common problems, identify good practices and coordinate both domestic and international policies.

Enlargement and Key Partners

OECD member countries agreed to open accession discussions with Colombia and Latvia in 2013, and with Costa Rica and Lithuania in 2015.

The Organisation is also reinforcing its engagement with its Key Partners – South Africa; Brazil, China, India and Indonesia.

Publishing

The OECD is one of the world's largest publishers in the fields of economics and public policy. [OECD publications](#) are a prime vehicle for disseminating the Organisation's intellectual output, both on paper and online.

Publications are available through the Online Information System (OLIS) for government officials, through OECD iLibrary for researchers and students in institutions, corporate, subscribed to our online library and through the Online Bookshop for individuals who wish to browse titles free-of-charge and to buy publications.

INSTRUCTIONS TO TENDERERS

ARTICLE 1 - PURPOSE AND OBJECT OF THE CALL FOR TENDERS

The OECD is issuing this Call for Tenders with a view to sourcing the design, development and pilot of an international assessment of children's early learning.

ARTICLE 2 - TERMS AND CONDITIONS OF THE CALL FOR TENDERS

2.1 Composition of the Call for Tenders

The documentation relating to the Call for Tenders includes the following parts:

- a) Instructions to Tenderers and their Annex;
- b) Terms of Reference and its Appendixes;
- c) Minimum General Conditions for OECD Contracts.

2.2 Tenders

All Tenders will be treated as contractually binding for the Tenderer and the Tenderer shall consequently issue in response to this Call for Tenders a Letter of Application dated and signed including all the provisions set out in clause 3.2 below.

2.3 Duration of Tender validity

Tenders shall remain valid for two hundred ten days (210) calendar days, as from the deadline for receipt of Tenders.

2.4 Additional information

Should any problems of interpretation arise in the course of drawing up the Tender documents, Tenderers may submit their questions to federica.darida@oecd.org and denis.elices-rejon@oecd.org , no later than seven (7) calendar days before the deadline for the receipt of Tenders. All Tenderers will be advised of the answers given to such questions.

INSTRUCTIONS TO TENDERERS

2.5 Acceptance and rejection of Tenders

There is no commitment on the part of the Organisation to accept any Tender or part thereof that is received in response to the Call for Tenders.

The OECD reserves the right:

- To accept Tenders with non-substantial defects
- To reject Tenders received after the deadline for receipt of Tenders, without indemnity or justification.

2.6 Modification or cancellation of Call for Tenders

The Organisation reserves the right to modify or cancel all or part of the Call for Tenders, should the need arise, without having to justify its actions and without such action conferring any right to compensation on Tenderers.

2.7 Partnerships.

Partnerships must jointly meet the administrative requirements set out in the Call for Tenders. Each partner must also meet full requirements individually.

2.8 Extension of the deadline for receipt of Tenders

The OECD reserves the right to extend the deadline for receipt of the Tenders. In that case, all the Tenderer's and Organisation's rights and duties and in particular Article 2.3 above will be subject to this new deadline.

2.9 Expenses

Tenders are not paid. No reimbursement of expenses related to the preparation of any Tender will be made by the OECD.

2.10 Confidentiality

The Call for Tenders and any further information communicated to the Tenderer or which come to his knowledge in the course of the Call for Tenders and the performance of the work, are confidential and are strictly dedicated to the purpose of the Call for Tenders. The OECD reserves the right to have all material returned at the end of the Call for Tenders process.

INSTRUCTIONS TO TENDERERS

ARTICLE 3 - PRESENTATION, SUBMISSION AND CONTENTS OF TENDERS

3.1 Tender presentation and conditions for submission

Tenders shall be entirely drafted **in English** and shall be **received** by the Organisation:

Before the deadline date: **Monday February 15, 2016 at 5.00 pm (Paris time)**

- In three (3) paper copies and one electronic version (e.g. USB Key):
- In a double anonymous envelope bearing the words:

**« NE PAS OUVRIR par le service courrier
Appel d'Offres n °100001420 »**

To the following address:

**OECD
EXD/PBF/CPG
To the attention of Federica DARIDA and Denis ELICES-REJON / Central Purchasing
Group
2 rue André Pascal
75775 Paris Cedex 16
FRANCE**

Tenders which are received after the deadline for receipt specified above as well as Tenders which do not fully comply with the Technical Specifications, may be rejected.

Tenders sent by e-mail or fax shall be systematically rejected even if they have also been sent in paper format (hard copy).

3.2 Contents of the Tender

- The Tender in **three (3) copies and one electronic version** (e.g. USB Key):
- A Letter of Application, signed by the Tenderer, confirming the following::
 - All the elements of the offer are contractually binding;
 - That the person signing the offer does have the authority to commit the Tenderer to a legally binding offer
 - That the Tenderer accepts all of the Minimum General Terms and Conditions without any modification. If there is an exception, please state the exception and the rationale for that exception.

INSTRUCTIONS TO TENDERERS

- That the Tenderer acknowledges and understands the terms of the Instructions to Tenderers and accepts to conform itself to those terms if it is selected to conduct the contract;
- That the Tenderer, or each of the partners in the case of a partnership, have fulfilled all its legal obligations with regards to tax declarations and payments in its home country and must supply all the requisite certificates to that effect;
- Moreover, the Tenderer shall provide, to the extent possible and where applicable, certificate(s) identifying the Tenderer, including its name, legal form, address, registration number or equivalent, date founded, areas of activity and number of employees ;
- The signed Declaration detailed in Annex to these Instructions to Tenderers.

Please note that the Tenderer, ***should it be shortlisted***, will be asked to provide the following:

- Any relevant existing agreements with intermediaries or third parties;
- Financial information for the last three (3) years
- Proof of completed legal obligations with regards to tax declarations and payments in its home country and all the requisite certificates to that effect;

3.2.2 Financial Conditions

Prices quoted must include everything necessary for the complete execution of an eventual contract (insurance, transport, guarantees). Charges for items essential to execution of the contract and not identified in the Tender will be borne by the Tenderer.

INSTRUCTIONS TO TENDERERS

ARTICLE 4 - INTERVIEWS

The Organisation reserves the right to organise interviews and request the Tenderers to specify the content of their Tenders.

ARTICLE 5 – SELECTION CRITERIA

Main criteria for Tenderer selections are detailed in section 6 of the Terms of Reference.

ARTICLE 6 - INFORMATION TO TENDERERS

All Tenderers will be informed, whenever possible, of the decision taken on their Tenders.

INSTRUCTIONS TO TENDERERS

Annex

Declaration Call for Tenders n° 100001420

As part of the offer in response to the OECD call for Tenders n° 100001420, the Tenderer (company or individual) declares on oath the following:

- That it is not bankrupt or being wound up, is not having its affairs administered by the courts, has not entered into an arrangement with creditors, has not suspended business activities, is not the subject of proceedings concerning those matters, and is not in any analogous situation arising from a similar procedure provided for in national legislation or regulations;
- That it has not been convicted of an offence concerning its professional conduct by a judgment which has the force of *res judicata*;
- That it has not been the subject of a judgment which has the force of *res judicata* for fraud, corruption, involvement in a criminal organisation or any other illegal activity detrimental to the interests or reputation of the OECD, its members or its donors;
- That it is not guilty of misrepresentation in supplying the information required as a condition of participation in this call for Tenders or fail to supply this information;
- That it is not subject to a conflict of interest;
- That its employees and any person involved in the execution of the work to be performed under the present Call for Tenders are regularly employed according to national laws to which it is subject and that it fully complies with laws and regulations in force in terms of social security and labor law;
- That it has not granted and will not grant, has not sought and will not seek, has not attempted and will not attempt to obtain, and has not accepted and will not accept any advantage, financial or in kind, to or from any party whatsoever, constituting an illegal practice or involving corruption, either directly or indirectly, as an incentive or reward relating to the award or the execution of the Contract.

I, the undersigned, on behalf of the company, understand and acknowledge that the OECD may decide not to award the contract to a Tenderer who is one of the situations indicated above. I further recognise that the Organisation may terminate for default any contract awarded to a Tenderer who during the award procedure had been guilty of misrepresentation in supplying, or fail to supply, the information requested above.

The .. / .. / ..

Signature

TERMS OF REFERENCE

SECTION 1 – INTRODUCTION

Overall goals

The OECD Secretariat is seeking an international contractor to design, develop and pilot an international study on children's early learning. The overall purpose of this study is to provide countries with a common language and framework to learn from each other and, ultimately, to improve children's early learning experiences. Countries interested in this study are particularly focused on improving equity of outcomes for disadvantaged children.

More specifically, countries'¹ objectives for this study are to better understand:

- Children's early learning and development in a broad range of domains, including social and emotional skills as well as cognitive skills
- The relationship between children's early learning and children's participation in early childhood education and care (ECEC)
- The role of contextual factors, including children's individual characteristics and their home backgrounds and experiences.

Countries wish to have a reliable and valid means to understand the variations in children's early learning and development, to see what is possible to achieve and to be able to monitor progress over time. Thus, the study has a system-level focus, and will not allow the identification of individual institutions, staff, children or parents. And while countries want comparable data, they wish to understand the relationship of this data to the contextual factors that underlie it.

Background

A proposal to investigate child outcomes across countries was raised by the OECD's ECEC Network in a priority-setting process in 2012. Following this, the Network developed *Common Understanding*, a document synthesizing ECEC policy and practice statements across a number of OECD countries². The document set out a number of principles relating to child outcomes that established a foundation for further work in this area. *Common Understandings* was published by the OECD in 2015.

The ECEC Network also provided input and oversight to a paper led by Dr Steven Barnett from Rutgers University on existing comprehensive measures of early child outcomes (attached as Appendix 1). The paper found that a number of reliable and valid measures of child outcomes have been developed and used, and that assessing child outcomes for a range of purposes was common in many countries.

During 2015, the OECD Secretariat developed a conceptual framework on early learning outcomes, in collaboration with interested countries. The framework sets out the inter-related key determinants of children's early learning. These are the child's individual characteristics, the child's home background and learning environment and the child's ECEC experiences. The latter comprises the child's participation

¹ Note that a Scoping Group of 16 countries has been working with the OECD to scope this study.

² *Common Understandings* is available on : <http://www.oecd.org/edu/school/ECEC-Network-Common-Understandings-on-Early-Learning-and-Development.pdf>

Terms of Reference

in ECEC as well as the type and nature of the ECEC provision. The final version of the conceptual framework is attached as Appendix 2.

In developing the conceptual framework, a set of principles were established, building on the guidance provided in *Common Understandings*. These principles set out parameters in which an early learning assessment should be developed, in terms of:

- Having policy relevance, ie enabling changes in policies and/or practices to be made
- Being practicable, ie able to be implemented
- Being reliable, valid and comparable across countries, languages, cultural contexts and over time
- Ensuring the well-being of children in the study is paramount in all decisions
- Limiting the burden on practitioners and parents, as well as on children
- Being affordable for a range of countries
- Taking a managed approach, to firstly establish a strong foundation for the assessment and then expanding from this base, if desired. For example, later cycles of the study may explore links to other OECD studies and to nationally-based assessments.

The objectives for this call for tender

The objective of this call for tender is to identify an organisation or consortium that – as an International Contractor – will design, develop, field-test, pilot and refine an international early learning study to provide countries with comparative data on children’s early learning outcomes.

Bidders are invited to consider and demonstrate how they would approach the design of an early learning study that meets countries’ objectives, whilst adhering to the principles set out above. In particular, bidders should consider how existing approaches and tools for early learning assessments could be incorporated with new design features to create a unique, fit-for-purpose and world-leading study. Bidders are also encouraged to consider the design of the study, including the assessment elements, as a whole. While countries have indicated preferences for the inclusion of individual domains, they also wish to ensure that the assessment as a whole is coherent and provides meaningful and insightful comparative information, that they can use to improve policies and practices in the field of early learning.

Note that the above timelines are indicative, and are dependent on a number of factors.

The preferred International Contractor will be:

- A recognised expert in large-scale early learning assessment
- Experienced in co-ordinating international projects
- Responsive to these terms of reference
- Innovative and pro-active in suggesting new approaches and alternatives in designing, developing and implementing this new study
- Able to work in conjunction with the OECD Secretariat in developing and refining the study
- Able to manage the complexity and challenges that designing and testing an international early learning assessment inevitably presents, and
- Flexible and willing to adjust its approach in response to findings from the field testing and pilot, and to feedback from the OECD Secretariat and the countries participating in this phase of the project.

Terms of Reference

The pilot study is intended to include between 3-6 countries. While a number of countries have shown a high level of interest in the study, and some of these have indicated their preference to be part of the pilot, some countries may not be able to formalise their participation in the pilot until the end of 2016. Thus, costings should be provided for carrying out the pilot in 3, 4 5, and 6 countries.

The organisation of the call for tender

There are six key tasks the International Contractor will undertake:

1. Design the overall shape of the study and develop assessment and other instruments accordingly
2. Establish protocols and procedures to achieve reliable, valid and comparable data
3. Develop manuals and other documentation to reflect the established protocols and procedures
4. Field test the assessment and other instruments in the 3-6 countries participating in the pilot
5. Pilot the study as a whole in these 3-6 countries
6. Analyse the findings from the pilot and amend instruments, protocols and procedures, and documentation, as appropriate.

Bidders are asked to bid for all tasks, and to consider partnership or sub-contracting arrangements as required.

Bidders should present their proposals separately for each of the six key tasks in this Statement of Work, provide a detailed response that describes how they will implement each of the requirements and tasks, and also submit a separate cost proposal for each of the key six tasks.

As the final shape of the study will continue to be refined through the development phase, bidders are encouraged to provide expert advice on possible options and the relative merits of these, in line with the stated objectives and principles for this study. Bidders may also propose alternative parameters to those outlined in these terms of reference, if such alternatives are likely to improve the quality of the data and/or provide efficiencies.

Bidders are expected to demonstrate their experience in co-ordinating international projects and the capacity to attract high quality scientific and policy evaluation and design expertise. It is important that bidders demonstrate that those who will design and develop instruments have a sufficient understanding of children's development, as well as home learning environments, ECEC systems and the social, cultural and educational environments in which children learn, including the educational systems and cultural contexts of a wide range of countries.

All data will be the property of the OECD, as will all test items, associated data and results. Bidders are asked to make clear their positions regarding the intellectual property rights implications of their proposed solutions and must make clear where third party rights are being used and therefore would not be able to be assigned to the OECD.

The International Contractor shall also ensure that the technology in the services offered remains compatible with technology advances. Bidders shall specify in the proposal if the proposed services will have any limitations in this regard.

Relationship with the new ECEC Staff Survey

The OECD is also developing a new ECEC staff survey, based on the Teaching and Learning International Survey (TALIS) in schools. Some countries have expressed interest in participating in both the ECEC staff survey and this new early learning study.

Terms of Reference

Bidders for the international early learning assessment should identify how they could build synergies between the ECEC staff survey and the early learning study. Such synergies should add greater value for countries that participate in both studies, in terms of both the depth of information and potential insights provided by undertaking both studies. In addition, bidders should highlight ways that countries can achieve greater efficiencies from participating in both studies. Bidders should list and quantify the potential optimizations and specify the expected benefits and savings.

SECTION TWO: RELATIVE ROLES AND STRUCTURES

This section sets out the expected roles of the International Contractor, the OECD Secretariat and the National Project Manager (NPM) in each participating country.

The International Contractor

The International Contractor will develop the overall design of the early learning study to meet countries' objectives, in line with the principles established for the project. Thus, the International Contractor will be responsible for designing the study as a whole, in addition to developing, testing, piloting and refining the assessment and other instruments in this study. As part of its management role, the International Contractor will prepare and maintain an overall project plan for each stage of the study, including implementation timelines for countries participating in the field testing and pilot.

Following the development of the overall design and project plan, the International Contractor will be responsible for developing assessment items in the domains agreed and for developing instruments for gaining information from parents and teachers. Thus, the International Contractor is responsible for assessment design and development, as well as for sampling requirements, manuals and other tools, training NPMs in assessment administration and for analysis of the findings.

The International Contractor shall be responsible for supporting and overseeing the preparations and implementation of the assessment in participating countries – from the first phases of translation, adaptation and field testing, to implementing the pilot. The International Contractor shall establish tools and procedures for effectively communicating with NPMs, for collecting and collating regular progress updates from NPMs, and for keeping the OECD Secretariat regularly updated on progress and issues arising.

The International Contractor shall be the main point of contact and communications with NPMs. The International Contractor shall specify and implement procedures that promote excellent communication with NPMs. The International Contractor will be expected to maintain a communication portal, where NPMs can communicate about tasks and where NPMs can find manuals, guidance and regularly updated information on progress with the survey.

The International Contractor shall call, organise and host meetings of NPMs. Provisions for meeting venues and facilities as well as for travel and compensation for experts, as required, should be included in bidders' proposals. No compensation of travel costs for NPMs or representatives from the OECD Secretariat should be included in the cost proposal. Participating countries will bear the costs of their NPMs participation in these meetings. Bidders shall propose the frequency of such meetings.

Bidders are encouraged to propose innovative and efficient approaches to meeting arrangements, eg through different meeting structures and an enhanced use of web-based communication tools.

The International Contractor shall also negotiate and resolve timeline amendments and minor disputes with NPMs, if they arise, in consultation with the OECD Secretariat.

The International Contractor will submit a monthly progress report to the OECD Secretariat covering all work included in the Statement of Work (Section 3). The International Contractor will also put in place procedures for monitoring risks, maintaining a regularly updated risk register and issues log, and provide regular updates on risks, issues and deviations from timelines as part of the monthly progress report to the OECD Secretariat.

Terms of Reference

The International Contractor will also establish mechanisms for submitting all study resources, documents, materials and databases to the OECD archive and for liaising with NPMs to ensure this is kept up-to-date.

The International Contractor shall nominate a senior person for the role of International Director. This person will be responsible for all activities and deliverables of the International Contractor and will work closely with the OECD Secretariat to ensure the success of this study. S/he will provide leadership for NPMs and should therefore have strong leadership and team-building skills. The person in this role should also have the credibility and experience to provide intellectual leadership amongst country representatives and other experts, and be able to identify and resolve technical issues. Bidders should name the person who will be carrying out this role and specify the percentage of time to be spent on the project by the International Director.

The International Contractor will be expected to present updates to meetings of the Early Learning Group as required. The Early Learning Group consists of countries who are participating in the pilot study and other countries who are interested in the development of the study. The Early Learning Group will provide advice and other input to the Secretariat on this study as it develops. Two face-to-face meetings of the Early Learning Group will be held each year, each over two to three days, in addition to shorter webinars and conference calls. The International Contractor will be responsible for covering travel, accommodation and other expenses for their own personnel who attend these meetings.

The International Contractor will also be expected to engage with a Technical Experts Group established by the Secretariat. The Technical Experts Group will provide input to the development of the assessment framework and instruments. The Technical Experts Group will help to ensure the study is internationally valid and reflects the cultural and curricular context of participating and interested countries.

Bidders should describe the type of expertise they would wish to have access to, in addition to the Technical Experts Group and the national experts represented on the Early Learning Group. In addition, bidders should specify the number of Technical Expert Group meetings they have included in their proposed budget, and should also describe how they would call on the expertise of group members outside the formal meetings. As already noted, bidders should consider the most efficient and cost-effective use of remote meetings as an alternative to physical meetings.

The OECD Secretariat

The OECD Secretariat is responsible for the overall management of the international early learning study. The Secretariat will work collaboratively with the Early Learning Group referred to above, to ensure countries' priorities and interests are reflected in the design and implementation of the study. Countries that participate in the field testing and pilot will do so through an agreement with the OECD Secretariat.

Formal reporting on the project is to the OECD's Education Policy Committee (EDPC).

The OECD Secretariat will participate actively during the development of all instruments, protocols and procedures, documents and reports and will approve all documents before they are provided to participating countries. This applies, in particular, to meeting documents, manuals and assessment materials.

The OECD Secretariat will also be responsible for:

Terms of Reference

- The active engagement of the Early Learning Group in the development and implementation of the study
- Keeping the EDPC, the PISA Governing Board and the ECEC Network regularly updated on progress and issues arising
- Ensuring a project management approach is agreed with the International Contractor and is applied to managing all aspects of the study
- Oversight of risks, issues and deviations from timelines, and ensuring risks and issues are regularly monitored and appropriately mitigated and managed
- Providing a central point for resolving any debates between the International Contractor and NPMs over responsibilities, workflow and timelines that have not been resolved through the processes of communication set up by the International Contractor
- Monitoring the budgets and milestones of the International Contractor and resolving budgetary or contractual issues
- Establishing and maintaining an archive of all project resources, documents, materials and databases
- Providing additional support to NPMs by attending NPM meetings, obtaining regular feedback from NPMs, and dealing with any queries or problems that cannot be resolved by the International Contractor.

The OECD Secretariat will also establish a Technical Experts Group, to provide the Secretariat, participating and interested countries and the International Contractor with specific technical expertise on specific areas of the study. The Secretariat will ask the International Contractor for advice on the areas of expertise that should be covered by the Technical Experts Group and for nominations for the Group.

The Secretariat may also appoint other experts, such as country-specific experts, following discussion with participating countries and the International Contractor. Such experts would provide advice on specific issues at particular points within the study.

To ensure the integrity of national samples, the OECD Secretariat will appoint a Sampling Referee for the pilot study. The Sampling Referee will inform participating countries and the International Contractor as early as possible of problems with sampling or response rates that may or will jeopardise countries' compliance with sampling guidelines, providing an explanation for the problems or concerns and, when possible, suggesting remedies for them.

The OECD Secretariat will arbitrate disagreements between participating countries and the Sampling Referee if such disagreements arise.

National Project Managers

Each pilot country will be required to appoint a National Project Manager (NPM), to implement the project at the national level subject to the procedures established by the International Contractor. The International Contractor shall develop a description of the role and profile of NPMs and specify the intended working relationships with NPMs.

NPMs will be the primary means of day-to-day contact between participating countries and the International Contractor for the implementation of the study. NPMs will play a vital role in ensuring that the study is a high quality project with results that can be verified and evaluated. The NPM will decide how to best facilitate the communication and co-ordination needed at the national level for implementing data collection responsibilities.

Terms of Reference

The NPM will be responsible for the translation of assessment items and other documents, if required, and any adaptation to the local context, supported by and following procedures set out by the International Contractor. The NPM will also be responsible for contracting and training in-country staff, such as assessors, and for liaising with School/Centre Co-ordinators.

The following table sets out the relative responsibilities of the International Contractor, the OECD Secretariat and NPMs.

Terms of Reference

	International Contractor	National Project Manager	OECD Secretariat
DEVELOP INSTRUMENTS			
Develop assessment framework	✓		
Design and develop instruments	✓		
ESTABLISH PROTOCOLS AND PROCEDURES			
Determine sampling requirements	✓		
Set assessment and data handling procedures	✓		
Translation to national languages and contexts		✓	
Verify translations	✓		
DEVELOP MANUALS			
Develop manuals	✓		
FIELD TEST INSTRUMENTS			
Oversee preparations with NPMs	✓		
Administer the field test		✓	
Analysis and validation of instruments	✓		
PILOT			
Develop pilot procedures	✓		
Communicate with schools/centres		✓	
Contract and train in-country staff		✓	
Administer the assessments		✓	
Code results		✓	
ANALYSE AND REFINE			
Develop reporting template			✓
Adapt reporting template to national contexts			✓
Translate reporting template		✓	
Data analysis	✓		
Brief countries			✓
Brief media and provide other communications, if required			✓
Refine assessment and other instruments	✓		
Amend manuals and procedures	✓		

SECTION 3 – STATEMENT OF WORK

TASK 1: DESIGN AND DEVELOP ASSESSMENT AND OTHER INSTRUMENTS

The International Contractor will develop an overall conceptual design for the early learning study. Based on this, the International Contractor will design an assessment framework to gather valid, reliable and comparative data on children's learning and development across countries in the range of domains set out in the conceptual design. The data will be captured at one point, thereby providing a snapshot of children's learning, rather than gathering data at two or more points over time.

The study will contextualise children's learning and skills in terms of each child's:

- ECEC experiences
- Home learning environment
- Individual characteristics.

The development of the assessment instruments should enable countries to establish and maintain trends over time, as well as to make in-depth comparisons with other countries. At the same time, the assessment design needs to be feasible and practical without overburdening national budgets and the time demands on children, their parents and teachers.

Age range for the assessment

Countries that have been involved in scoping this study have expressed a clear preference for an age-based rather than stage-based target population, to support valid international comparisons of children's early learning outcomes. Countries have agreed that the assessment should occur at the point where "nearly all" children are in some form of education or care provision. Countries have agreed on an age band between 4.5 to 5.5 years, although not all countries will be able to assess children across this entire age band. For example, some countries will only be able to reliably sample children from 5-5.5 years.

Thus, the International Contractor will need to propose a methodology using age adjustments and weightings to ensure comparability and validity of the data across countries. Implications of the recommended approach for sampling should also be identified.

Domains to be assessed

The domains to be assessed should represent a balance of both cognitive and social and emotional skills that, as a package, will provide coherent and reliable insights into children's early learning. The domains selected should be those that are malleable in the early years, including in ECEC environments. Six possible domains have been identified, based on an analysis of early skills that are predictive of positive life outcomes³ and through consultation with interested countries, as follows:

- Self-regulation
- Oral language/emergent literacy
- Mathematics/numeracy

³ An analysis of predictive early learning skills has been carried out by the UCL Institute of Education for this study.

Terms of Reference

- Executive function
- Locus of control
- Social skills

Each of the domains is outlined below.

Self-regulation

Self-regulation generally encompasses self-control, grit, self-management and conscientiousness. These abilities enable children to persist in achieving goals and to regulate their behaviour. The latter manifests through inhibiting impulsive behaviours and delaying gratification (Mischel et al., 1989). As well as achieving tasks, children with such abilities are more able to operate effectively in groups than children with poor behaviour regulation (OECD, 2015).

Oral language/emergent literacy

Oral language skills comprise those skills required to speak, listen and understand, and include vocabulary knowledge. There are several domains within oral language, including:

- the sounds produced while speaking (phonemes)
- the rules a given language requires to construct sentences (syntax), and
- the understanding that concepts have meaning (semantics).

Emergent literacy refers to children's knowledge of print, letters and sounds, which will help them to learn to decode and read for meaning, building upon oral language skills. For this part of the domain, it will be important to reflect country differences in the ages at which children are introduced to reading and writing.

Numeracy

Numeracy is the ability to reason and apply simple numerical concepts. It comprises the ability to identify and understand numbers as well as computational skills, ie the ability to count and to perform simple arithmetical operations such as addition, subtraction, multiplication, division, and compare numerical magnitudes. In early numeracy, children are detecting patterns and beginning to understand that things can be measured.

Executive Function

Executive function focuses on the ability of children to regulate attention, including controlling reactions to new stimuli. The capacity to regulate attention is understood as a developmental precursor for the broader domain of self-regulation (Barkley, 1997). Executive function additionally provides information on working memory and planning, which are also associated with later academic development (Bull et al., 2008; Nesbitt et al., 2015; Sasser et al., 2015).

Terms of Reference

Self-awareness/Locus of control

Self-awareness refers to children's own beliefs about whether they possess the ability to complete tasks, and encompasses aspects such as self-esteem, self-confidence, self-efficacy and locus of control (John and De Fruyt, 2015). Locus of control refers to whether a person believes their own performance is based on external factors, (ie perceiving that the action of other people or luck determines an outcome), or internal factors (ie perceiving that they have control over an outcome). This attribution style relates to having either a "fixed mindset" (believing that capabilities are inborn and unchangeable) or a "growth mindset" (believing that capabilities, including intelligence can be developed and increased) each of which leads to different behaviours and achievement (Dweck, 2008).

Locus of control has not commonly been included in assessments of early learning, but the analysis undertaken by the UCL Institute of Education referred to above has highlighted it as highly predictive of later life outcomes. The International Contractor will be invited to advise on the usefulness and feasibility of including locus of control in the study, alongside other recommended domains.

Social skills

Social skills include pro-social behaviour, agreeableness, sociability and empathy. Social skills are those skills involved in interacting with others and maintaining positive relationships with others. In particular, collaboration requires the ability to take the perspective of another, to demonstrate prosocial behaviour (ie showing kindness, sharing, co-operation, and respect for others), agreeableness and empathy. Ways of approaching others have been conceptualised in terms of extraversion, assertiveness or leadership, sociability, popularity and likability, as well as the capability of developing trust in others and the ability to communicate effectively (Schoon et al., forthcoming).

The domain on social skills could, for example, focus on:

- pro-social behaviours, including co-operation
- empathy
- trust.

Assessment Framework

Once the package of domains is finalised, including clear identification of synergies and independent impacts, the International Contractor will develop an assessment framework for each domain, which will be the conceptual underpinning for that assessment and will inform the nature of the information to be collected, the outcomes to be measured and the approach to developing assessment items and related tools such as questionnaires.

Following input from the Early Learning Group and approval by the OECD Secretariat of the assessment framework, the International Contractor will develop assessment items and instruments for each domain. The International Contractor will draw on the Technical Experts Group, as well as other relevant experts as required.

Assessment could be undertaken through a range of methods, from multiple observations by independent, trained assessors through to practitioner and parent questionnaires. Options may be provided on possible

Terms of Reference

means of assessment, with advice on the relative merits of each in terms of reliability, validity and comparability, and the impacts on participants in the study and on overall cost.

Where possible, bidders should set out the assessment instruments they would be likely to use, develop or revise for each domain, how the assessment would likely be carried out, and how much time the assessment of each domain would be likely to take. Where existing instruments are proposed, the rationale for doing so should be clear, along with the countries and languages these instruments have been or are being used in and any evidence on efficacy should also be highlighted. Where revisions or new instruments are proposed, the proposal should set out the rationale for this and how the appropriateness and applicability of such instruments will be assessed.

Bidders should describe their proposed processes for assessment development and investigation of validity, including the suitability, cultural appropriateness and reliability of assessment items during the course of their development. This could include, for example, the use of laboratory investigations or focus groups.

Contextual information

Contextual information provides insights into the relationships between children's learning and development and important demographic, social, economic and educational variables.

The International Contractor will develop a framework for the collection of contextual information. This will include information relating to the individual child, the child's home learning environment and ECEC experiences, as well as relevant institutional and system-level data.

Bidders are encouraged to consider innovative approaches to collecting contextual information about the children who are being assessed, their home learning environments and their ECEC experiences from several sources. Bidders are also asked to suggest how such information can be collected in ways that are culturally sensitive, suitable for use in a wide range of countries and that do not present a burden for parents or school/centre practitioners.

The International Contractor will be asked to look into ways of cross-nationally validating the data collected in the participating countries and economies. Bidders should describe how they would ensure the cross-national comparability and validity of the instruments to gather contextual information, including suitability and cultural appropriateness.

Participating countries may modify the wording or format of items, or add national components to the questionnaires used to collect background or explanatory information from parents or school/centre staff. Many contextual items need to be specified and agreed in their nationally specific format, such as descriptions of education levels and home languages. The International Contractor shall provide guidance for such issues to NPMs and shall set up procedures for working with NPMs to approve the content of these national adaptations before they are included in the pilot for a particular participating country.

ECEC Participation

Information on the child's ECEC history should include:

- The age of entry to ECEC
- The intensity of the child's ECEC participation, eg part-time, full-time
- The duration of ECEC participation, eg number of months

Terms of Reference

- Continuity, ie the number and duration of breaks between periods of participation, and the number of different providers/settings the child has attended, and
- Type of provision, such as the type of setting and pedagogical approach.

There are many differences in ECEC approaches both within and across countries. Bidders should suggest how these differences will be accounted for, whilst enabling countries to have reliable, valid and comparable data on the variations in children's early learning and development relating to children's ECEC experiences.

Home learning environment

Aspects of the home learning environment that are of particular interest are:

- The learning and development-related activities parents undertake with their children and the frequency of these activities. Such activities are likely to include some or all of the following: reading books, singing, telling stories, playing games, doing puzzles and doing art and crafts
- The extent to which parents are actively engaged in their child's ECEC experiences, where the child is participating in ECEC. This includes the extent to which parents and ECEC staff share information and strategies to support the child's learning and development.

Individual characteristics

The aspects of children's individual characteristics to be collected relate to each child's:

- Gender
- Ethnicity
- Linguistic background
- Migration history and status
- Socio-economic and family background
- Disabilities and special learning needs.

The definition of socio-economic status should be aligned to the PISA index of Economic, Social and Cultural Status (ESCS). The PISA student background questionnaire includes items on parents' occupations and education levels, in addition to indicators of household wealth. Bidders are also asked to consider the feasibility of gathering information on parents' mental health.

TASK 2: ESTABLISH PROTOCOLS AND PROCEDURES TO ACHIEVE RELIABLE, VALID AND COMPARABLE DATA

The study is to be designed to provide comparable data across a wide range of countries. Considerable effort will need to be applied to achieve cultural and linguistic breadth and balance in assessment and other materials. Stringent quality-assurance mechanisms will need to be applied in the assessment design, translation, sampling and data collection.

Thus, the International Contractor will develop procedures for the appropriate administration of the early learning assessment, to ensure it is administered under the same conditions in all settings in all countries, and that the results are comparable.

Terms of Reference

The International Contractor will also develop technical standards for the implementation of the study, specifying the quality requirements in terms of sampling, translation and translation verification, assessment administration, quality monitoring, coding, data entry and data submission, and release and exclusion of data.

A number of countries have expressed interest in the early learning study, who will not participate in the field testing or pilot. It is therefore important that the International Contractor designs the study in a way that ensures it is applicable to a wide group of countries, to enable such countries to join the study at a later stage if they choose to.

The International Contractor shall develop and implement survey operations, procedures and related aspects of quality control, including the development of assessment administration procedures and the training of all necessary and relevant country representatives in these procedures, eg NPMs, assessment administrators. The International Contractor will develop all related training materials and procedures in consultation with the Technical Expert Group and the OECD Secretariat. All training materials shall be developed in English.

The International Contractor will be responsible for monitoring that all NPMs are following established technical and other standards. Bidders should describe in detail how they would do this and what requirements they would place on NPMs to monitor and provide assurance that all quality standards are being met.

Translation

The International Contractor will be responsible for developing all materials in English. The International Contractor will advise on translatability issues and prepare translation guidelines for NPMs. The International Contractor will also develop and propose technical standards regarding the translation of assessment instruments, questionnaires and manuals, for consideration by the OECD Secretariat.

English will be the language of communication between the International Contractor and NPMs. Functionalities and materials that are only to be used by NPMs need not be translated by the NPM. Examples of such materials are the sampling tool described below and the manual for NPMs.

NPMs will be responsible for translations following procedures set up by the International Contractor. The International Contractor must work with NPMs to ensure that the translations reflect the language as used in each participating economy/country and are of a quality that will ensure the cross-national comparability of the assessments.

The International Contractor is responsible for verifying the linguistic quality of the translation. The translation verification is carried out to ensure that the translation reflects the language as used in the participating country and is of a quality that ensures the cross-national comparability of the assessment. While the translation is the responsibility of the NPM, the translation verification will be the responsibility of the International Contractor.

Thus, the International Contractor will put in place procedures for managing translation verification and for resolving any disputes with NPMs. Bidders should describe their approach to this aspect of quality control, such as through the use of a Translation Referee or by another means.

Sampling

Terms of Reference

Countries wish to base the assessment on a sample of “nearly all” children. In many countries access to a sample of all children may be difficult and costly. Thus, countries’ preference is that only children enrolled in educational or other institutions will be assessed, at a point when all or nearly all children are enrolled.

Appendix 3 sets out participation rates in some form of education or care setting for children for a range of countries.

Bidders should propose how a comparable and valid sample of 4.5 to 5.5 year old children can be obtained. Children within the target age group are in different types of settings in different countries, eg in ISCED 0.2 or ISCED 1 institutions. As seen in Appendix 3, in some countries it may be relatively easy to access children aged 4.5 to 5.0 while in others access is easier from age 5, as noted earlier.

The sampling design for an international early learning assessment also requires particular attention because of the fragmented nature of the provision for the selected age group, e.g. there are a multitude of ECEC providers in some countries, while in other countries a significant share of children in the selected age tranche may already be enrolled in early primary schooling. This should be taken into consideration when proposed a sampling approach and, as described in Task 1, an age adjustment may be needed to ensure comparability. Lastly, to correct for a potential selection bias, bidders should provide a proposal to account for the fact that in countries where there is participation in both ECEC and in primary school across the target age group, participation rates vary across these settings.

As in many other international educational surveys, such as the Programme for International Student Assessment (PISA) and the Teaching and Learning International Survey (TALIS), the international early learning study will take a multi-stage sampling approach, ie first taking a representative sample of institutions providing education and/or care to children to the defined age group and then sampling children within each of the selected institutions.

A sampling framework is required that will deliver representative samples of the target population/s. Samples must be designed to maximise sampling efficiency for child-level estimates. However, they should also permit the linkage of child-level data (eg learning assessment results, individual characteristics) with institutional-level and parent-level variables (eg the home environment, ECEC experiences) that are collected through background questionnaires. Notably the importance of linking children’s learning to participation in ECEC has been highlighted by interested countries.

Bidders should propose the ideal sample sizes for the instruments being proposed. Samples must be designed to maximise sampling efficiency. However, they should also permit the linkage of data on children’s learning with the contextual variables relating to the child’s individual characteristics, home learning environment and ECEC experiences, as indicated above.

Principles

The bidder’s proposal should demonstrate how the following principles will be upheld with regard to the sampling approach, drawing on those set out in the conceptual framework in Appendix 2:

- Reliability, validity and comparability
- Efficiency and cost-effectiveness.

The bidder should also propose considerations with regard to a potential stratification of institutions in the sampling frame, to be carried out prior to sampling. This may allow the proposal to:

Terms of Reference

- Improve the efficiency of the sample design
- Improve the reliability of survey estimates
- Apply different sample designs, such as disproportionate sample allocations, to specific groups of institutions, eg in specific states, provinces or regions
- Ensure that all parts of the target population are included in the sample
- Ensure that specific groups of the target population are adequately represented in the sample.

Interest in analysing sub-groups of the population may relate to those children from particular socio-economic groups or backgrounds, or those in particular settings. Thus, these interests may affect the sample sizes recommended by the International Contractor.

The International Contractor will

- Prepare draft and final versions of sampling plans for the field tests and the pilot study. Sampling plans should specify methods and standards for decisions regarding inclusion/exclusion, mechanisms for assessing the adequacy of participating countries' sample frames and for assuring the adequate demographic representation of children and schools/centres
- Advise on sampling standards and develop quality control procedures for ensuring the sampling standards are met. In co-operation with the Sampling Referee, establish procedures for dealing with samples that do not meet the predetermined sampling standards
- Work with NPMs to define the target population and draw the samples, and advise NPMs on their sample design.

TASK 3: DEVELOP MANUALS

The International Contractor shall develop manuals that outline the protocols and procedures referred to in the previous section. The manuals shall be geared to the following target groups:

- *National Project Manager Manual*
The manual will as a minimum include an introduction to the early learning study, protocols for communication between the NPM and the International Contractor, procedures for preparing participation in the programme such as translation, adaptation and field testing, and the implementation of the pilot. The manual should be written in a language and with a level of detail that takes into account that no NPMs will have had experience with this programme before. The International Contractor will also be required to provide training to NPMs on assessment administration, based on the manual.
- *School/Centre Co-ordinator Manual*
The manual shall give the school or centre all the information necessary to prepare for the assessment appropriately. The manual shall as a minimum include an introduction to the study, a description of the School/Centre Co-ordinator's role, guidelines for the preparation of a list of eligible children from which children will be sampled and a child tracking list to track children's participation on the day of the assessment; and detailed instructions on how the assessment will be administered and overseen.
- *Quality Monitor Manual*

Terms of Reference

The manual shall give Quality Monitors, employed by the NPM, the information necessary to oversee that procedures for the appropriate use of the assessment are being followed and that the school or centre's preparations and follow-up to the assessment are done appropriately. The manual shall also describe how to report back to the NPM on any deviations observed or feedback received from the school or centre.

The International Contractor shall develop source versions of the manuals in English. The NPM in each participating country will be responsible for translating and adapting the manuals to the local language and context. The source versions shall include clear indications of the elements that NPMs are expected to adapt to the local context.

TASK 4: FIELD TEST THE ASSESSMENT AND OTHER INSTRUMENTS

The assessment and other instruments will be tested in the field, to investigate the feasibility and operationalization of the assessment approaches, to ensure robust and reliable instruments in each participating country.

Where assessment instruments are new or revised, prior to field testing, these instruments will undergo small-scale item trialling. Bidders who are proposing the development of new or revised instruments should describe their intended approach to trialling and specify the extent of likely new item development and item trialling.

All instruments will be subject to field testing in the countries participating in this pilot study. The International Contractor shall support each country in testing the items. The recruitment of schools and/or centres for the field testing will be the responsibility of the NPMs, while the International Contractor will be responsible for developing the overall procedures and design.

Following testing in the field, the International Contractor will analyse the field testing results, validate the instrument parameters and prepare the final version of the assessment and other instruments.

TASK 5: PILOT THE STUDY IN 3-6 COUNTRIES

The International Contractor shall support NPMs with the implementation of the pilot. As part of this management role, the International Contractor shall provide and maintain tools for NPMs to track progress with the implementation of the tasks involved with the pilot in each country and to keep track of any potential problems with countries' abilities to meet project timelines or technical standards.

The NPMs will follow procedures and timelines defined by the International Contractor. NPMs will be required to:

- Recruit schools/centres, in line with sampling procedures determined by the International Contractor
- Negotiate assessment windows in consultation with schools/centres, the International Contractor and the national education authorities
- Develop a test administration schedule for each school/centre. The school/centre-specific administration schedule will take into account the assessment window/s established, the school/centre's preferred dates for the assessment, and the availability of human and logistical resources in the school/centre

Terms of Reference

- Obtain a list of eligible children from the designated School/Centre Co-ordinators and, in co-ordination with the School/Centre Co-ordinators, to identify children who may be eligible to take the assessment but who will be excluded at the school/centre level for specific reasons, ie as established in the standards and procedures. It is important that child exclusions are registered and documented
- Co-ordinate with the schools/centres to allow parents an “opt-out” opportunity from the assessment for those children who are selected as part of the school/centre sample. If necessary, the NPMs may need to provide additional information to school/centre practitioners and parents to build an understanding of the importance of children participating in the assessment
- Communicate the list of eligible children to the International Contractor, following the procedures set out by the International Contractor, and receive a list of sampled children from the International Contractor
- Ensure correct procedures for administering the assessment at the school/centre level
- Ensure that assessment materials are kept secure and school/centre information is kept confidential
- Identify, contract, train and co-ordinate the activities of Quality Monitors. The Quality Monitors will oversee that the established assessment and other related procedures are followed, collect feedback that can be used to improve the procedures, and resolve any technical or other issues that may occur on the day of the assessments
- Collect questionnaires and other requested information.

TASK 6: ANALYSE FINDINGS AND AMEND INSTRUMENTS AND DOCUMENTATION, AS AGREED

The International Contractor will clean all collected data, conduct analyses on the pilot study data, and provide a fully documented database with a set of basic indicators or their components, which will allow the OECD Secretariat to conduct its own further analyses. Bidders are asked to indicate the types of checks that will be carried out on the data, and the mechanisms that will be put in place to ensure that checks are carried out by NPMs, as required.

The International Contractor shall also use the data collected during the pilot study to conduct analyses to identify problems in the implementation of the administrative procedures, investigate methodologies of data analyses, investigate the properties of measurement instruments and carry out other analyses that may be necessary to refine the instruments and survey procedures.

Participating countries wish to obtain a thorough understanding of the variations in learning between different groups of children and the contextual factors that may influence these differences. Bearing this in mind, bidders are asked to give some preliminary ideas for the types of analysis that they see the potential to conduct and the new policy insights that could potentially be provided from such analysis. Bidders are also asked to describe the techniques they see as most promising for developing these policy insights.

The International Contractor will develop an analysis and reporting plan, which will guide the OECD Secretariat in preparing and designing the reporting on the pilot study. The plan should summarise and explain the types of analyses that can and should be conducted to address the objectives of this study, and also how the data can best be presented and reported. For example, this will include designing and providing basic descriptive tables following a standardised format that the Secretariat specifies. The International Contractor will provide statistical and technical support to the OECD Secretariat during the

Terms of Reference

development of the report. The International Contractor will review the report and drafts of the report for technical consistency and coherence.

The International Contractor will be responsible for leading the development of a Technical Report. The Technical Report should serve the needs and answer likely questions from a range of audiences, from sophisticated survey and data experts to those without expertise in this area. Bidders should suggest how they would ensure the report serves the needs of all users.

The International Contractor will review and revise the technical standards, procedures and manuals, following the implementation of the pilot, as agreed with the OECD Secretariat.

SECTION 4 – INDICATIVE TIMELINE

The OECD Secretariat envisages that the work to develop and/or modify assessment instruments and test these in the 3-6 countries that will participate in the pilot study will be undertaken during 2016 and early 2017. The pilot of the instruments as a whole is intended to be undertaken at the end of 2017 or during the first half of 2018. The aim is to complete the work covered by these terms of reference by the end of 2018.

Bidders are requested to include a detailed timeline for the delivery of the different tasks and elements of each task, included in these terms of reference. Bidders are encouraged to split the work into phases and specify when each phase will be complete.

Final details of the schedule and work plan for the study will be determined following discussion with the International Contractor and the OECD Secretariat. A detailed project plan with key milestones and deliverables will be completed by the successful International Contractor within an agreed timeline.

SECTION 5 – COSTING MODELS AND FINANCIAL OFFER

The OECD requests bidders to provide financial offers that recognise the options for the assessments and information gathering. Bidders are encouraged to propose and cost different models of responding to this brief. It is important to both meet the objectives of this study and provide countries with options that are affordable for a wide range of countries and that represent value-for-money.

Financial offers must be presented in EUR and detailed according to the following table.

	2016	2017	2018
Task 1: Design and develop assessment and other instruments			
Task 2: Set protocols to achieve valid, reliable, comparable data			
Task 3: Develop manuals and procedures			
Task 4: Field test assessment and other instruments in the 3-6 countries intending to participate in the pilot			
Task 5: Pilot the study as a whole in 3-6 countries			
Task 6: Analyse findings and amend instruments and documentation, as agreed.			

The budget for Tasks 4 and 5 must be provided for 3, 4, 5 and 6 countries participating to the Pilot.

Terms of Reference

Budget for options (if any) proposed by the bidder must be quoted separately.

Costs should be given separately for each of the tasks, and broken down to clearly show the expected budget for management, meetings, consultants and contractors, and miscellaneous costs including administrative support and materials. The budget information should include a breakdown of individual staff costs and roles and relevant workload, as per the below table:

Item	# Units (Days)	Cost (€) per Unit (Daily rate)	Total (€)	Level of seniority of team member
Task 1				
Staff 1				
Staff 2				
Staff 3				
Other Costs (please specify)				
Total Other Costs			€	
Total Cost			€	

The budgetary worksheet must be submitted in a separate paper file from the rest of the response to this Call for Tenders and as a separate electronic file on the USB key or CD. For the electronic file, it must be submitted in .doc or .xls format.

SECTION 6 – EVALUATION CRITERIA

The evaluation criteria which will be used to assess the technical and financial merits of the bids received are as below:

40% **Technical quality**, which includes but is not necessarily restricted to:

1. The extent to which the proposal reflects an understanding of and responds to the intent and direction in the Terms of Reference, i.e. clear, convincing and feasible proposals for each of the tasks listed in the Statement of Work, in line with the study objectives;
2. The technical quality of the project design and implementation plan, to ensure that the study will produce data which are reliable, valid and comparable across countries, and provide valuable policy insights based on the results.

30% **Expertise and capacity**, which includes but is not necessarily restricted to:

1. Experience and capacity in early learning assessment design and implementation in an international context;
2. Proven capacity in developing collaborative relationships that promote consensus-building;

Terms of Reference

3. Proven capacity in project and budget management.

30% **Financial proposal.** The project must be affordable and represent value-for-money for a wide range of countries, as well as deliver reliable, valid and comparable data. Bids will be evaluated on the proposed pricing and the justification provided for the costs associated with each component and set of activities, including alternative and optional activities in the proposal.

References

- Barkley, R. (1997), “Behavioural inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD”, *Psychological Bulletin*, Vol. 121, No. 1, pp. 65-94.
- Bull, R., K. Andrews Espy and S. Wiebe (2008), “Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years”, *Developmental neuropsychology*, Vol. 33, No. 3, pp. 205-228.
- Dweck, C. (2008), “Can personality be changed? The role of beliefs in personality and change”, *Current Directions in Psychological Science*, Vol. 17, No. 6, pp. 391-394.
- John, O. and F. De Fruyt (2015), *Social and emotional skills framework for the longitudinal study of skills development in cities*, OECD, Paris.
- Mischel, W., Y. Shoda, and M. Rodriguez (1989), “Delay of gratification in children”, *Science*, Vol. 244, No. 4907, pp. 933-938.
- Nesbitt, K., L. Baker-Ward, and M. Willoughby (2013), “Executive function mediates socio-economic and racial differences in early academic achievement”, *Early Childhood Research Quarterly*, Vol. 28, No. 4, pp. 774-783.
- OECD (2015), *Skills for social progress, The power of social and emotional skills*, OECD, Paris, http://www.oecd-ilibrary.org/education/skills-for-social-progress_9789264226159-en.
- Sasser, T., K. Bierman, and B. Heinrichs (2015), “Executive functioning and school adjustment: The mediational role of pre-kindergarten learning-related behaviors”, *Early Childhood Research Quarterly*, Vol. 30, pp. 70-79.
- Schoon I., Nasim B., Sehmi R. and R. Cook (forthcoming), *The Impact of Early Life Skills on Later Outcomes, Final Report*.

APPENDIX ONE – COMPREHENSIVE MEASURES OF CHILD OUTCOMES IN EARLY YEARS

COMPREHENSIVE MEASURES OF CHILD OUTCOMES IN EARLY YEARS: REPORT TO THE OECD

STEVE BARNETT, SHANNON AYERS, & JESSICA FRANCIS

3. In this report we provide basic information to inform decision-making regarding the assessment of young children's learning, development, and well-being for national and international data collections designed to inform Early Childhood Education and Care (ECEC) policies. Our primary focus is on the pre-primary years with an emphasis on assessments that are relevant to a broader age range including older children. Given the large number of assessments available, this report begins with a broad overview and then considers specific examples of the various approaches to illustrate strengths and weaknesses rather than conducting an exhaustive review. Several much broader reviews with exhaustive compendia are already available that can be consulted. These include major publications from the U.S. National Academy of Sciences (Snow & Van Hemmel, 2008) and the World Bank (Fernald, Kariger, Engle, & Raikes, 2009).

1. Why is early childhood assessment important?

4. Assessments of young children can provide information about Learning, Development, and Well-Being (LDWB) that is useful to teacher, parents, and others. For teachers assessment can be a tool that informs the care and education they provide to children. Parents often wish to be informed about the progress and wellbeing of their children, less to inform the specifics of their interactions than to be assured that their children are doing well and that the arrangements they have made for their care and education are in the child's best interests. Program administrators can use child assessment data to explore the effectiveness of program design and supports for teachers including professional development. With respect to public policy, there are several valuable uses. Screening, and where indicated, diagnostic assessments conducted on a large scale can identify disabilities and other developmental problems so that children's special needs can be addressed as early as possible. Nationally representative descriptions of children's LDWB and how this varies geographically and with children's family backgrounds as well as with the characteristics of children's ECEC experiences can inform a wide range of public policies to support children and families, keeping in mind that drawing valid causal conclusions regarding public policies and programs is a complex process and imposes strong demands on research design and analysis, not just on assessments.

5. As nations increase their public and private investments to support the care and education of young children, it is to be expected that they will want information about the contributions of these investments to the lives of young children. In particular, there is increased concern about how specific

public policies affect children before they enter primary school. This desire to establish cause and effect and to estimate the magnitude of benefits to children's LDWB increases the technical demands on assessment (discussed below). In addition, causal attributions require more than simply children's describing LDWB over time, it requires rigorous research methodologies that warrant strong causal inferences. Historically, relatively little information of this type has been collected by public agencies prior to age 8, well after entry to primary school in many countries.

1.1 Use and concerns

6. Broadly speaking, the use of assessments can be described as formative or summative. Formative assessment is the use of assessment to inform teaching with some definitions going so far as to equate formative assessment with scaffolding. Formative evaluation is internal and takes place during the educational experience. It looks forward in a process that is responsive to the needs of the learner. Summative assessment is the use of assessment to judge progress or attainment relative to a standard. Summative assessment of the performance of a child looks backwards and may be used to judge the contributions of a teacher or program to child progress. Summative assessment generally is external in its orientation. Summative assessments may be used to inform professional development and other supports for teachers and programs, but they also may be used to make "high stakes" decisions including to sanction or reward teachers, schools, and to inform decisions about public programs and policies. In addition, summative assessments are commonly used to make high stakes about individual children including the provision of additional supports (e.g., special education services and services for immigrant children who have limited proficiency in the local language) and opportunities (e.g., programs for gifted children), as well as to determine whether a child should enter primary school at the typical age or delay entry. The last use is quite controversial and may be viewed as indicating a lack of supports and individualization in the first year of primary school.

7. As it is the use of an assessment that is formative or summative rather than the assessment instrument itself, the same instrument can be used for summative or formative purposes. Confusion can arise because of instruments have been designed so as to be particularly useful for formative or summative purposes, and sometimes the instruments themselves are referred to as formative or summative measures. In addition, there is a tendency to think of qualitative assessments and teacher observations as formative tools. However, the Early Years Foundation Stage Profile (EYFS Profile), discussed in the next section, is an example of an observation-based assessment that is used for primarily summative purposes. The EYFS Profile also provides an example of how the distinction between formative and summative use can become less clear when looked at from a longer term perspective. Even though the assessment is not used to inform teaching in the child's current stage, it is used to inform the child's education at the next stage and to inform changes in policy.

8. Despite the widespread use of assessment, there is widespread concern regarding potential negative consequences. Among the greatest concerns are (1) narrowing of ECEC to focus on what is most easily measured; (2) misuse of assessments for high stakes decisions about children, teachers and programs; and, (3) excessive burdens on children and teachers from time consuming assessments. These concerns have been greatest for direct tests used for summative purposes, but they may arise with any type of assessments regardless of the use for which it was developed. For example, screening tests are sometimes misused to make high stakes decisions about children rather than to refer them for additional assessment. Screening tests also are sometimes used to collect data on a large scale to inform policy because they are quick to administer and so impose minimal costs on everyone involved. However, this should be done in full recognition that screening tests often are designed to err on the side of over identifying problems and may measure better at the lower end than at the higher end of the range of abilities or skills.

9. Concerns about negative impacts of assessment on learning and teaching and misuse by policy makers are lessened when assessment is conducted with broad observational measures embedded in the educational process for formative purposes. As we discuss in more detail in later sections, teachers may document in detail children's interests, dispositions, learning, development, and well-being as a tool to assist them in providing the best care and education for each child. Yet, even the kinds of data teachers collect for these purposes can be turned to other, summative purposes. Moreover, because the broader and more detailed such assessments become, the greater the time burdens they may impose on teachers.

1.2 Current policy and practice

10. Today, assessment of children's learning and development in the years before the age of 5 is common in OECD countries. Typically teachers conduct such assessments as an integral part of their teaching; most often these are not standardized tests, but ratings, observations and collections of children's work. Less often, teachers formally assess children's well-being, but teachers frequently make judgments about each child's well-being (e.g., happiness, self-actualization, and friendships). Although considered good ECEC practice, it is uncommon for these ECEC assessments by teachers to be required by law. Most often whether, how and when this assessment is conducted is up to the discretion of ECEC providers. As it is good practice, it may be encouraged by public policy guidance. However, some countries (or states in federal systems) require assessment by law, and a few of these specify the assessment to be used and when it is to be used.

11. In a recent survey, OECD countries varied in the extent to which they reported that assessments were used for monitoring purposes in ECEC programs. Many reported that assessments were used as monitoring for summative and formative purposes, with formative use more common. Ireland, Italy, Korea, the Netherlands, and New Zealand reported that they did not use assessments for monitoring purposes. However, to some extent reported differences among countries appear to reflect differences in their interpretation of the questions as well as differences in practice. Generally, the use of assessments by teachers--particularly rating scales, checklists, portfolios, and storytelling--to improve practice is ubiquitous across countries. Much less common is the use of standardized tests for such purposes or the use of standardized tests for any purposes. Standardized tests most often are used to assess language and literacy, motor skills and physical development. Observations and ratings most often assessed children's LDWB very broadly across many domains. The use of assessments for external evaluation of program performance was rarely reported outside of the Americas.

12. Assessments of young children also are collected in national longitudinal and panel studies and, less often, national evaluations of specific ECEC programs. Such studies assess only a sample of young children rather than every child in an age cohort. Typically national samples include children from all ECEC arrangements including those who are only at home with family members. Data from these studies can be used to inform policy makes and the public of the status of children's LDWB and how it is changing over time. Its usefulness for these purposes increases with the frequency with which it is collected.

13. Some public policies regarding assessment are noteworthy as indications of the extent of international variation. France offers teachers the option to use a national test at age 5 for the purpose of better understanding the children they serve and for the teachers' exclusive use. Austria has compulsory language tests 15 months before entry to primary school (Stevens & Dworkin, 2014, p. 71). The purpose of these tests is to ensure that children who do not have adequate German language proficiency receive additional assistance with language development in kindergarten. The tests used differ from one state to another. Germany mandates language tests in kindergarten for similar reasons, but these tests differ by

Terms of Reference

state. Finland has asked children to evaluate their ECEC programs through photos, drawings, and evaluation forms as well as interviews (though these last were conducted only with 48 children).

14. England mandates assessments when a child is between 2- and 3-years-old and at the end of the school year when they turn 5. These assessments are based on teachers' observations. The age 5 assessment is conducted using a rating scale (with a brief narrative), the Early Years Foundation Stage Profile (EYFS Profile). This Profile was recently redesigned, and a new version was introduced in 2012. The Profile provides a broad assessment of child development that is aligned with the EYFS standards. Information is obtained from parents and from teacher observations. The EYFS Profile provides information to parents regarding their child's progress, to the current teacher for use in transition discussions between teachers, and to the teacher who will receive the child in the first year of key stage 1 of primary school for individualized educational planning. In addition, the EYFS Profile is used to construct "an accurate national data set relating to levels of child development at the end of the EYFS which can be used to monitor changes in levels of children's development and their readiness for the next phase of their education both nationally and locally" (Standards and Testing Agency, 2013, p.7). The resulting performance tables are not published at the school-level.

15. Another common use of assessment among OCED and other countries is large scale screening followed by clinical diagnostic assessments to identify disabilities and other developmental problems (including vision and hearing problems) that would benefit from early treatment. However, we did not locate systematic international data on national policies regarding screening and diagnosis of disabilities, delays, and other developmental problems (including hearing and vision limitations). In some countries screening takes place quite early while in others it is not required until well after entry to primary school.

1.3 Illustrative variations within the United States

16. As education policy in the United States varies greatly amongst the 50 states, a brief review of such policies provides insights into the range of different policies that might be adopted. Of the 40 states that offer publicly funded preschool education programs (typically at age 4), the vast majority require the use of some assessments, though not necessarily specifying the assessment or even the type of assessment. Most often this assessment is to be used for formative purposes by teachers, but most states also seek to use this information to inform teacher professional development. A few states require assessments for high-stakes decisions about children (e.g., kindergarten entry) or for summative purposes including the evaluation of teacher and program performance for sanction or reward. Such states may specify a specific assessment to be used with every child enrolled. State policies regarding preschool assessment are summarized in Table 1 below (Schilder and Carolan, 2014).

17. Most states have or will soon adopt Kindergarten Entry Assessments (KEAs) that measure learning and development when children enter kindergarten (the first year of primary school) after turning age 5. The use of these assessments also varies considerably by states. Some states intend these assessments to provide a broad baseline measure that describes children as they enter school. This information would be used by teachers to inform their practice, but also could be aggregated to inform policy makers about the needs of young children and to assess growth between entry at age 5 to kindergarten and the next time the state mandates uniform assessment of every child, typically at the end of third grade. KEAs often have a "whole child" perspective and are not narrowly academic. However, some state KEAs focus primarily on early literacy and, sometimes, a few other academic domains such as mathematics. A few states plan to use these assessments to judge the educational effectiveness of individual ECEC providers and for this purpose the KEA may be aligned with an earlier assessment in the preschool years.

Table 1: U.S.A. state uses of assessments in Pre-K

How Pre-K Assessment Data Are Used By the States	Number of state programs
Guide teacher training, professional development, or technical assistance	35
Track child and program level outcomes over time	34
Make adjustments to curricula	32
Provide a measure of kindergarten readiness	17
Make changes to state policies regarding the preschool program	16
Make decisions regarding a child's enrolment in kindergarten	6
Identify programs for corrective action or sanctions	5
Make funding decisions about programs or grantees	5
Evaluate teacher performance	2

2. Overview: Deciding what and how to assess.

18. From the perspective of obtaining national or international data that can be used to inform policy rather than practice, there are key criteria to be used in deciding what and how to assess. These criteria are as follows:

1. Measure what matters. What aspects of LDWB are important and of concern to policy makers and the public?
2. Measure well. To be useful, measures of what matters must be valid, reliable, fair, and age and developmentally appropriate.
3. Assessments must be practical and affordable. The younger the child, the more difficult it is to accurately assess their LDWB. The broader and deeper the assessment the higher the cost. In addition some aspects of LDWB are more difficult and expensive to assess. Time demands on children, teachers, parents and others can be substantial (opportunity costs such as lost time from teaching), and the costs of professionals specifically hired (and trained) to administer assessments or interviews may be high as well.
4. Results of assessments should be comparable within and across countries and over time.

2.1 Measuring what matters.

19. Children's LDWB encompasses virtually every possible outcome of ECEC including children's happiness and life-satisfaction, habits and dispositions, attitudes and beliefs, cognitive abilities, social abilities, emotional development, physical development, health, and nutritional status. Such a broad view is consistent with the early childhood field's emphasis on attending the needs of the whole child. In addition, one might add measurement of the extent to which a child's rights are respected, for example, the right of children to have a voice or active role in determining the activities in which they are engaged in ECEC. This could be viewed as a means to producing outcomes for the child (for example, life satisfaction and attitudes toward society and schools). However, it could be viewed as an additional category.

20. Both common values and research indicate the importance of comprehensive measures. In most, perhaps all, countries the goals of ECEC are to support the development and well-being of the whole child. This is evident in the Convention on the Rights of the Child (Melton, 2011). From a child's rights perspective we may include "opportunities to express personal agency and creativity, feeling able to contribute, love and care for others, to take on responsibilities and fulfil roles, to identify with personal and community activities, and to share in collective celebrations (Woodhead & Brooker, 2008, p.4). It is also evident in a U.S. National Academy of Sciences report on the science of early childhood development (Shonkoff & Phillips, 2000) which recognized the value of:

(1) the development of curiosity, self-direction, and persistence in learning situations; (2) the ability to cooperate, demonstrate caring, and resolve conflict with peers; and (3) the capacity to experience the enhanced motivation associated with feeling competent and loved (p.5).

21. Note that we have not described any of these domains or their measures as "outcomes." The use of the term "outcomes" raises the question: Outcomes of what? Children's learning, development, and well-being are affected by all of their experiences at home and with family more generally, in ECEC arrangements, and in the community as well as of their personal attributes. Drawing valid inferences about the specific influence of ECEC experiences and the policies that shape them is much more complex than simply looking at correlations between ECEC and child LDWB measures in a cross-section or longitudinally. One might call for randomized trials, and it is sometimes possible to conduct these with special data collections or in such a way that they can use data that would have been collected anyway. However, randomized trials are not always possible or ideal. It is much more likely that comparisons of the impacts of ECEC and ECEC policies within and across countries will be conducted using complex statistical models that are more successful in producing valid inferences when there are assessments at multiple time points (at least one "pre-test") and when the assessments are accurate and precise. These statistical methods also benefit from linked information on each child's family, home experiences, and ECEC experiences.

22. What should be assessed does depend on the purposes for which an assessment will be used. If policy makers wish to evaluate differences in ECEC quality and services, these may be expected to influence some aspects of learning and development more than others. For example, if the vast majority of young children is healthy and has good motor development, and these are carefully monitored by health professionals then ECEC programs may not much affect these domains. In this case, there may be little reason for educators to assess them. If there is a strong concern that children's rights to engagement and active decision making are not adequately respected, then this aspect of wellbeing may be an important focus of assessment.

2.2 Measuring well: Desirable features of assessments

23. To be useful assessments should be valid and reliable. Assessments also should be fair. In early childhood there is particular concern that assessments be age and developmentally appropriate. This applies equally to all types of assessments, performance assessments as well as tests, qualitative as well as quantitative.

24. Validity is a fundamental criterion for selecting instruments to measure LDBW. The *Standards for Educational and Psychological Testing* state, “Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed use of tests” (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999, p. 9). A valid instrument—whether an observation, interview, questionnaire, or test—should measure what it purports to measure (Williams & Monge, 2001). Validity refers to the appropriateness, meaningfulness and usefulness of the specific inferences made from an instrument so that it is always judged in the context of the purpose for which those inferences are made (Borg, Gall, & Gall, 1989). In assessing validity, what we wish to know is the extent to which interpretations of a measure hold across persons and contexts.

25. In essence, validity is established by producing and evaluating evidence on how well an assessment represents the construct it purports to measure (Messick, 1995). Validity depends on the extent to which an assessment represents the entire construct (i.e., it is not enough that items be from the appropriate domain, they must be fully representative of it). Validity also requires that a measure not include irrelevant items (as, for example, when language demands obscure the demonstration of math or social skills). In other words, an assessment can be invalid because it is too narrow and shallow or because it is too broad. An assessment also can be invalid because accurate representation does not generalize across populations and contexts.

26. There are multiple types of evidence that help to establish construct validity. These include assessments of content by experts, structural evaluation, comparison against a criterion, and prediction. The extent to which experts concur that an assessment fully covers the key dimensions of the construct being measured and does not tap irrelevant areas is sometimes referred to as face validity. Validity is also judged based on the structure of the assessment. Do patterns of results across items conform to theoretical expectations regarding the underlying concepts? Criterion validity can be assessed by examining patterns of performance across ages and concurrent correlations with other assessments of the same construct. A high degree of correlation with an instrument that has well-established validity provides evidence supporting the validity of the target assessment. At the same time a valid measure should not be highly correlated with a measure that is believed to measure a completely different construct. Other approaches include estimating the extent to which the assessment predicts current or subsequent performance in “real life” that is contingent on what is measured.

27. Assuring validity for many assessments is not simply a matter of design, but also of assuring that procedures are appropriate for individual children. An obvious issue occurs when a child’s home language differs from that of the assessment. Another is when a child has a disability, and this is most easily understood with respect to vision and hearing impairments. With respect to both issues accommodations often must be made to the child in order to maintain the validity of an assessment.

28. Reliability is the extent to which an assessment produces stable or consistent results because it produces little random error in its results (Creswell, 2008). A reliable assessment produces the same or highly similar results for a child on different occasions (assuming only a brief interval between assessments) and with different assessors (e.g., one teacher would not rate the same child differently from another teacher). A reliable assessment is also robust with respect to the circumstances of the assessment.

Terms of Reference

29. Reliability can be improved through several means. Optimizing the length or detail of an assessment is one way to increase reliability. The more items, or samples, obtained the less random error affects the results, unless, for example, a longer “test” results in fatigue or distraction for the child or assessor. Another is to construct items and their scoring so as to maximize clarity and minimize uncertainty or misunderstandings. Minimizing the influence of incidental factors in the environment or assessment circumstances and subjective (idiosyncratic) interpretation also increase reliability as does guidance and training for the assessors.

30. Multiple approaches are available to evaluate reliability. One of the most common is examining internal consistency, or how the items (or samples) in the assessment relate to one another. Historically, reliability as judged by internal consistency has been assessed using Chronbach’s Alpha, though recently this approach has been challenged and others recommended as more appropriate (Yang & Green, 2011). All of these approaches produce reliability coefficients (a measure of correlation among items). In general, tests that have a reliability of .80 or higher can be considered as sufficiently reliable for most research purposes (Borg, Gall, & Gall, 1989). However, reliability coefficients should be judged carefully since the value adequacy depends on the phenomenon studied (Hancock & Mueller, 2010). Values of .90 have been recommended for assessments used for high stakes decisions about individuals (Yang & Green, 2011).

31. Other common measures of reliability are the correlations of repeated assessments of the same child by the same assessor and inter-rater agreement of different assessors. Inter-rater agreement also may be assessed as criterion-related observer reliability, which is the extent to which a trained observer’s scores agree with those of an expert observer (Borg, Gall, & Gall, 1989). It is important because it declares that the trained observer understands the variables measured in the instrument with the same efficacy as an expert observer. Again, there are norms with respect to the extent of agreement required and this depends on the use with the highest levels of agreement required when use relates to an individual child. A high level of reliability is important not just when use is summative but also when used to inform individualized education of a child.

32. Fairness refers to the ways in which assessments are used rather than a property of assessments per se. In addition, it is socially defined rather than scientifically defined. In our view, fairness does depend on validity and reliability because for an assessment’s use to be considered fair most would agree that the assessment should be free of bias (e.g., with respect to gender, family background, or national origin) and that random error should not be higher for some types of children than others (at least at the same age). However, even a valid and reliable assessment can be applied in ways that are not fair.

33. One concern in the early childhood field is that assessments developed for older children not be pushed down to younger children when they are neither age nor developmentally appropriate. This concern arises, in part, because of the much greater availability of assessments for older children than for younger children. As demands grow to assess young children on a broader set of domains for which fewer assessments are available, for example, creativity and subjective wellbeing this temptation to use inappropriate assessments only increases. The problem can be avoided by limiting assessments to those with substantial evidence of validity and reliability, which depend on instruments being age and developmentally appropriate.

2.3 Practical Issues: Feasibility and cost

34. In addition to meeting the criteria for validity, reliability, and fairness, a desirable assessment or set of assessments is feasible and affordable. Otherwise, it will not be used or, if used, will create unintended negative consequences. Depending on how information is collected assessments impose costs

on children, parents and teachers as respondents or observers. More detailed and comprehensive assessments impose higher costs on respondents and assessors. In addition there are costs for purchasing assessment tools and training those who administer and use them. If assessment results are made available to people other than those collecting the data there are costs of the systems for storing and sharing data, as well. Finally, the younger the child, the more practical difficulty there is in obtaining information without placing unrealistic demands on the capacities of the child or assessor. As discussed earlier, one of the costs of excessive demands is deterioration in the quality (reliability) of the information obtained. In addition, imposing (unreimbursed) costs on teachers, parents, and others will increase nonresponse rates.

35. The costs to purchase assessments or the tools for their use are minor compared to the actual costs of training assessors and administering assessments. Yet, policy makers sometimes ignore the later and act as if there is no cost for administration if teachers (who are already paid) conduct the assessments. This assumption seems especially likely when assessment is formative and integrated into teaching. However, there is always an opportunity cost, and for teachers this can be quite high. The cost of time spent in classrooms collecting, recording, and reviewing assessment information is best measured by the value of the activities that teachers forego as a result--this can be other forms of planning, but is likely to include direct caring and education of children. Similarly, it should be recognized that parents' time is not "free," and while it is desirable to obtain multiple perspectives, requesting that they provide information imposes opportunity costs on them, as well. Ultimately, the time costs imposed on parents and teachers may result in costs to children by decreasing the time they have to interact with children.

36. Different types of measures not only have different costs, but differ in who bear those costs. For example, brief direct tests or parent interviews (to obtain ratings) impose some costs on children and parents, but could have substantial costs for specially trained assessors who administer the instruments. Direct tests administered by teachers might be brief individually, but require substantial time if obtained for every child in a setting. Depending on the nature of the test it might be perceived by children as enjoyable (e.g., a game) or stressful. On the other hand, portfolios or rating scales completed by teachers may be collected unobtrusively without interfering with the children's activities and requiring no parent time or outside staff. However, teacher assessments may require many hours observing children and recording the results rather than interacting children in ways that directly enhance their wellbeing, learning, and development.

2.4 Comparability

37. For OECD and national policy-related purposes the results of assessments must be broadly comparable over populations and time to be useful. Most countries now have substantial numbers of children from different linguistic, cultural and national backgrounds, and such diversity is by design in international comparisons. Policy questions often span substantial periods of time so assessments should be comparable over a lengthy period. In addition, policymakers often have an interest in continuity and change over the life course so useful assessments will be comparable, at least to some extent, across ages. In this last instance, it is not necessarily the case that scores or ratings would be strictly comparable. What it means for a child to be competent at math or to be highly creative or what social behaviour is considered to be appropriate varies with age. An assessment might ask how many times a young child has a physical conflict in a day as a pre-schooler and in a year as a teenager. However, it is desirable that results on an early childhood assessment of a given construct predict results on assessments of the same construct in the primary and secondary years.

38. Differences in language and culture raise concerns regarding comparability. International studies must confront the problem that languages (and cultures) are not comparable in every respect. Of course, this is an issue that is commonly dealt with in international assessments including the IEA Preprimary

Study conducted of children's abilities at age 7. Often this is addressed by expert translation and back-translation evaluated by professional opinion and statistical assessment of differential item functioning supported by qualitative approaches (Ercikan, 2002; Benitez & Padilla, 2014). Even within a single country language issues arise. Variations in language within countries may be longstanding or the result of recent migration. Children may be monolingual or multilingual. With multilingual children, it must be decided whether the goal of language assessment is to measure the child's knowledge and proficiency in a single language or across all of the languages used by the child.

3. Approaches to assessment.

39. Information on children's LDWB can be collected through a variety of methods, both quantitative and qualitative. Assessments vary in the extent to which they are standardized and in the source (or sources) of their information. Information on children can be obtained directly from children or from those who observe them, most often parents and teachers or other adult caregivers. It may even be obtained from other children (nominations of friends or evaluations of peers to assess relationship status and social skills).

3.1 Tests

40. In education, the first type of assessment that comes to mind for many people is standardized tests. They are widely used to assess cognitive abilities, particularly to assess academic achievement in specific content areas. However, this approach has been used to assess a wide range of cognitive abilities. They were developed to increase the reliability, validity, and especially to increase the fairness of assessments by reducing assessor (particularly teacher) bias. Standardization refers not just to the instrument itself, but also to the process of its administration. It aims to reduce random fluctuations in the circumstances and procedures, and to eliminate systematic biases by the assessor through variations in procedures as well as subjective judgment. Tests may be group or individually administered. As our focus is only assessment prior to primary school we review only individually administered tests; group administered tests are not recommended for children in this age range due to inadequate reliability and validity.

41. Games, including digital games, may be viewed as a type of test. They may be explicitly designed to assess specific knowledge and skills. Their administration and scoring can be more or less standardized. They can be administered "on demand" as are tests generally or children may play them in the ordinary course of their activities. Thus, games as assessments, can share some of the characteristics of authentic or performance assessments, which are discussed in the next section, below.

3.2 Performance assessments and qualitative interviews

42. Another broad type of assessment with which most early childhood professionals are familiar is performance, or authentic, assessment for which observation of children in their everyday activities is the primary basis for data collection (Dunphy, 2008). These assessments typically are embedded in teaching and data are collected continuously during the year and as part of ordinary activities. Documentation can include notes, and observation records, artifacts, art, dictation and children's writing, photographs, and video and audio recordings. Conversations with children and clinical interviews (in-depth, open-ended, and highly sensitive to individual interviewed) are related qualitative methods that may be used to collect information when the phenomena of interest are difficult to observe.

43. The documentation obtained can be collected and organized in portfolios for each child. As developed by Reggio Emilia and other constructivist approaches, representation is a means to involve

children in self-assessment as well as for sharing information among teachers and parents (Dunphy, 2008). The results of such assessments may be communicated in highly qualitative form in learning stories and other narratives or quantified in ratings or scores. Narrative approaches seek to maintain the whole child perspective and recognize the inter-relatedness of children's dispositions, habits, skills, and knowledge as well as the importance of context for understanding children's LDWB.

44. The procedures for conducting and reporting or scoring performance assessments vary from highly standardized to completely unstandardized. Clinical interviews are by design highly individualized and unstandardized, though their methodology is standardized. Such interviews could be considered a separate type of data collection on their own, and have been used to assess children's perceptions of their own decision-making and influence in ECEC (Sheridan, & Pramling Samuelsson, 2001). However, this seems to have been done more often to characterize classrooms or programs than to represent the wellbeing of an individual child.

45. Some performance assessment systems are linked to specific curricula and provide tools and detailed procedures for data collection and scoring based on rubrics. Specific training in the assessment system is part of the professional development for learning to use the curriculum. Some other performance assessments are more general while others are highly specific, but so much a part of an emergent and developing approach to curriculum that they are no more standardized than the curriculum.

3.3 Checklists and rating scales.

46. A third type of assessment frequently used is the checklist or rating scale. Performance assessments can be scored using a checklist or rating scale (and accompanying rubric) either at one point in time or recorded periodically over a year. However, in this section we refer to measures do not necessarily require continuous data collection over time (and are summative in use). Instead, parents, teachers, or other adults rely on their general knowledge of the child or a brief current observation to answer questions about the child's capabilities, personality, dispositions, behaviour, or other characteristics. Such assessments may be standardized in the sense that the precise form and order of the questions has been devised based on research and are not be varied.

3.4 Time diaries

47. Time diaries collect data about children's activities including information about the types of activity, duration of each activity, the place of each activity, and who else was engaged with child as well as what else may have been going on at the same time. They provide a unique and very detailed, approach to assessing children's engaged capacities and wellbeing. Multiple techniques are available including: "beeper studies" where an activity is recorded when prompted, observation, written short recall, and telephone interview short recall. For example, a telephone survey might obtain a 24-hour record by asking a parent: "beginning at midnight yesterday what did your child do?" Basically, this is one long open-ended question with many prompts. Such methods typically are not used with young children. However, parents have been asked to report in telephone surveys for infants and young children, and teachers have been asked to complete written diaries for children in ECEC centers (Barnett & Boyce, 1995; Hofferth & Sandberg, 2001; Rossbach, 1988).

3.5 The different types of informants

48. Information on children's LDWB is obtained not only with different types of techniques, but also from different types of informants. These include parents and other (informal) caregivers, pre-

Terms of Reference

primary and primary teachers (we include here all those responsible for the care and education of children in formal settings including some family home care), and health professionals. In addition, children themselves are key informants and can be active participants in their assessment. There are advantages and disadvantages for each informant when assessing young children's LDWB. Informants may provide information directly or professionals specifically trained to administer an assessment may be employed to obtain information, typically from children and parents and other caregivers.

49. Parent and other caregivers are valuable informants because of the intimate knowledge they acquire of a child due to their relationship and the time they spend with the child. However, if caregivers are asked to provide ratings relative to an implicit standard or expectation (for example regarding learning, development, relationship quality, life satisfaction or happiness) they may differ greatly from one socio-economic environment and culture to another regarding what is typical or normative (Ertem et al., 2008). Caregivers also tend to provide socially desirable answers. Despite these disadvantages, caregivers' information about children can be valuable and nationally representative information can be readily obtained through household surveys. Some checklists and rating scales are designed to be relatively robust with respect to variations among parents.

50. Teachers in preschool or school settings often provide valuable insights into children's LDWB, though they can only report on those children who attend ECEC programs. Teachers make good informants because they tend to spend a great deal of time with the children and have working knowledge of and/or training in learning and development. However, teachers vary considerably in their preparation and training. This can be expected to greatly affect their ability to evaluate children's LDWB, especially with performance assessments. The less standardized and more qualitative an assessment, the more the quality of the results (validity, reliability and fairness) depends on the teacher's knowledge and skills regarding both LDWB and assessment. For many instruments, specialized training of the teacher (or other assessor) may be required.

51. The ratio of students to teachers varies considerably and can be fairly high in some countries. Differences in ratios can affect how well teachers know each child and how much time is required of the teachers if asked to assess all of the children for which they are responsible. For these reasons, ratios may be expected to affect the quality of assessment and reduce the reliability (though not necessarily the validity) of assessments in some countries or subpopulations within a country compared to others.

52. Health professionals have advantages as informants of children's development because of their understanding of how children progress through development and in some instances the health services may be the only professional services available to young children. However, for some health professionals, monitoring child development can be a new concept (Ertem et al., 2008). Also, the familiarity with the child and the level expertise possessed by health professionals can vary by socio-economic context. As with teachers, this can affect the reliability and, perhaps, validity, of the assessment.

53. Children are always, in a sense, the basic source of the information on their LDWB. Often, this is indirect and mediated by others. However, young children can provide direct responses in tests, other direct assessments, and interviews. They can be asked to provide ratings. The younger the child, the greater the difficulty of obtaining direct information that is valid and reliable.

4. Critical review of exemplars for comprehensive assessment young children

54. The vast number of options available to assess the LDWB of young children presents a challenge to any review. The number of domains, approaches, and purposes has called forth many

Terms of Reference

different assessments. Fortunately, others have provided exhaustive reviews of the available assessments. Prominent among them are efforts by the World Bank and U.S. National Academy of Sciences (Fernald, Kariger, Engle, & Raikes, 2009; Snow & Van Hemmel, 2008). Additional exhaustive compendia have been developed for the U.S. Department of Health and Human Services (Berry, Bridges, & Zaslow, n.d.) and the state of Washington (Slentz, Early, & McKenna, 2008). Also, useful is a more focused critical review that addresses key issues in both tests and authentic assessment (Atkins-Burnett, 2007). Currently, UNESCO is developing a Holistic Early Childhood Development Index and has conducted a review of early childhood development and wellbeing indicators to support that project (Tinajero & Loizillon, 2012). A very broad review of early childhood development indicators from an international perspective is provided by Frongillo, Tofail, Hamadani, Warren, and Mehrin (2014).

55. The existing compendia describe the instruments with respect the domains covered, ages at which they are appropriate, methods of administration, strengths and weaknesses, time for administration, and cost. For any specific assessment one may wish to consider they provide a key resource to which readers of this paper can refer. Our purpose here is to consider key illustrations of different approaches that are better known and may be considered guides to the most appropriate possibilities for an international assessment in OECD countries.

56. A list of the instruments reviewed here and the domains that they cover is presented in Table 2, below. In addition, a summary of the information collected on each assessment is presented in a separate set of 3 matrices. Detailed narrative descriptions and evaluations of each follow.

Table 2. Domains covered by exemplar assessments (domains not represented on the table are absent because they were not included in any of the assessments reviewed)

	Physical	Social/ Emotional	Cognitive	Communi- cation and Language	Executive Function	Approaches to Learning	Arts/ Creativity
Zambian Child Assessment Test	<input type="checkbox"/> (fine motor)	<input type="checkbox"/> (parent report)	<input type="checkbox"/> (information processing, nonverbal reasoning)	<input type="checkbox"/>	<input type="checkbox"/>		
NIH Measures	<input type="checkbox"/>	<input type="checkbox"/> (psychological well-being, stress, social relationships, negative affect)	<input type="checkbox"/> (EF, attention, memory, language, processing)	<input type="checkbox"/>	<input type="checkbox"/>		
Brigance Early Childhood Screens	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			
Denver II	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			
Griffiths Mental Development Scales Extended revised	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			
Mullen Scales of early learning	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> (visual reception)	<input type="checkbox"/>			
Schedule of Growing Skills	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			
Hong Kong Early Development Scale	<input type="checkbox"/> (gross & fine motor, physical fitness)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			
ELS	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> (math, science, literacy)	<input type="checkbox"/>	<input type="checkbox"/> (self regulation)		

Terms of Reference

	Physical	Social/ Emotional	Cognitive	Communi- cation and Language	Executive Function	Approaches to Learning	Arts/ Creativity
International Performance Indicators in Primary Schools (iPIPS)	<input type="checkbox"/> <input type="checkbox"/> (parent survey)	<input type="checkbox"/> <input type="checkbox"/> (teacher rating)	<input type="checkbox"/> (math, literacy)	<input type="checkbox"/>	<input type="checkbox"/>		
Work Sampling System	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> (literacy, math, science)	<input type="checkbox"/>			<input type="checkbox"/>
Teaching Strategies GOLD	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> (self regulation)	<input type="checkbox"/> (attends & engages, persists, solves problems, curiosity, motivation, flexibility & inventive thinking)	<input type="checkbox"/>
High Scope Child Observation Record	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> (math, literacy, science, social studies)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
EDI	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> (social competence & emotional maturity)	<input type="checkbox"/> <input type="checkbox"/> (literacy & numeracy)	<input type="checkbox"/>	<input type="checkbox"/> (aggressive behavior, hyperactivity, inattentive behavior)	<input type="checkbox"/> (independence & adjustment)	
Kindergarten Entrance Inventory for Connecticut	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> (numeracy, literacy)	<input type="checkbox"/>			<input type="checkbox"/>
Ages & Stages Questionnaire	<input type="checkbox"/> <input type="checkbox"/> (gross & fine motor)	<input type="checkbox"/>		<input type="checkbox"/>		<input type="checkbox"/> (problem solving)	
Parents' Evaluation of Developmental Status	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			
Battelle Developmental Inventory	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> (attention and memory)		
Child Development Inventory	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			
Early Years Foundation Stage (EYFS)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> (attention)	<input type="checkbox"/> (engagement motivation thinking critically)	<input type="checkbox"/>

4.1 The Zambian Child Assessment Test (ZamCAT; 2012) provides an example of a broad assessment that was constructed by adapting a range of existing instruments each of which is designed to measure specific domains and using variety of methods, but primarily from a one to one direct assessment (testing) perspective.

Purpose: The ZamCAT is a population measure administered to preschool children along with the standard population-based household survey. The ZamCAT is available at http://developingchild.harvard.edu/activities/global_initiative/zambian_project/

Age: Preschool

Format and administration: The ZamCAT followed a mixed approach in development of the tasks on the assessment. Several of the tasks included in the ZamCAT are existing assessments with some adaptations

Terms of Reference

where appropriate; while other tasks are newly developed. The ZamCAT is administered in partnership with the population-based household survey to preschool children by a trained examiner.

The ZamCAT assesses 7 domains of child development by blending existing measures with newly developed tasks. Each component of the assessment is described here. First, ZamCAT evaluates **fine motor skills** through two tasks. In the first task, the child is asked to copy letters, numbers, and a triangle using a pencil (taken from the Development Assessment in Zambia; Ettling, et al., 2006). The second is a newly developed timed activity where the child is asked to string beads on a shoelace, place beans into a cup, unbutton and button a shirt, and play a variation of the traditional game nsolo.

Both receptive and expressive **language development** are assessed in the ZamCAT. *Receptive vocabulary* refers to words that a child can comprehend and respond to, even if the child cannot produce those words. The ZamCAT examines receptive vocabulary with 30 items heavily adapted from the Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 2006). These were modified to be culturally and linguistically appropriate for Zambian children. The authors of the assessment note that scores on this component of the ZamCAT cannot be used to compare to the PPVT (Fink, Matafwali, Moucheraud, and Zuilkowski, 2010).

Expressive vocabulary refers to words that a child can express or produce. The ZamCAT examines this by posing two questions to the child 1) Can you tell me about something exciting that happened to you? 2) Can you tell me about the people you live with at home? These two questions were taken from previous research by Matafwali (2010). The responses are scored 0 (non-responsive) to 5 (multiple-sentence answer using correct grammar). The authors note that this sub-test was used particularly well across languages (Fink, Matafwali, Moucheraud, and Zuilkowski, 2010).

Nonverbal Reasoning is assessed on the ZamCAT with two tests. The first is a newly developed Object Pattern Reasoning (OPR) which uses patterns with concrete items. Here the child is asked to complete patterning sequences. The second test of nonverbal reasoning is taken from the NESPSY (Kirkman, et al., 1998) which is a series of neuropsychological tests. The NESPSY Block Test is used in the ZamCAT as an assessment that measures the child's ability to capture, analyze, and replicate abstract forms. In this test the child is asked to assemble blocks in reproduction of a pictured design.

Information Processing is assessed on the ZamCAT with the Rapid Automatized Naming (RAN; Dencla & Rudel, 1976) task. This task has the child look at pictures, colors, letters, or numbers and then name them as quickly as possible. However, for the ZamCAT only the picture subtest of this assessment was used. Photos for this test include chair, tree, bicycle, duck, and scissors.

Letter naming was included in the ZamCAT to examine children's preparedness for early literacy. Children were given two minutes to name letters shown in random order on a piece of paper.

Executive Functioning is assessed on the ZamCAT through two tasks. Attention is examined through the Pencil Tapping test (Brooker, Okello, et al., 2010). This is where children need to remember and apply the "rules" of the game (when to tap with the pencil) and the task is made more difficult by also providing the child another small task to divide his or her attention. Executive functioning is also examined by assessing a child's delayed gratification (impulse control). Previous assessments used either candy or a wrapped gift to measure this component of executive functioning. The ZamCAT offers one candy immediately or two candies if the child waits until the assessor is done talking to parents to receive his or her treat. Several issues arose with this task. For instance, some parents didn't allow candy, children were reluctant to take candy from strangers, or they lost candy to older siblings so assessors needed to give candy to all of the family.

Terms of Reference

Socio-emotional Development is measured by parent report on the ZamCAT. This is a series of 20 questions to capture parents' overall perceptions of development. The responses to the questions regarding if the child displayed the behavior are never, sometimes, usually, always.

Task Orientation is rated by the child evaluator. The evaluator rates children on their attitude and performance during the child assessment tasks. The rating scale measures executive function, compliance, and attention as rated by the child evaluator.

Developmental domains covered: This assessment examines nonverbal reasoning, receptive and expressive language, fine motor skills, information processing, socio-emotional development, task orientation, and executive function through direct assessments of young children. Additionally, this survey instrument includes an extensive questionnaire regarding the mother's health and health care during pregnancy and the child's health during the first few years of life.

Time Required: The total battery - including the child assessment and the questions asked to caregivers - takes between 90 and 120 minutes; the child assessment itself takes between 30 and 45 minutes on average, but varies quite a bit depending on how easily the child manages to do the tasks (G. Fink, personal communication, August 12, 2014).

Training and materials: Training for Administration is extensive. The assessors in previous studies have participated in training for 5 days, which could be considered on the short side for this type of assessment. The researchers tried to give feedback through supervisors on a daily basis during field work. It is recommended that 2 weeks of training with extensive field trials be implemented as a more rigorous approach to training. Training usually is divided into 5 parts: 1) Getting familiar with the tool: objectives, concepts, procedures; 2) Introduction to tool administration: rules, guidelines, and practical issues, including a mock assessment by trainers; 3) Within group practicing - 2-3 full assessments of other trainers; 4) Translation: group interviewers grouped by language to review the translation of all instructions and items; 5) Supervised field tests which are full assessment of 3-5 children, partially supervised (G. Fink, personal communication, August 12, 2014).

Technical (psychometric) properties: The ZamCAT reports reliability information using the Chronbach's Alpha. The Cronbach's Alpha coefficients reported for fine motor, receptive language, pattern reasoning, pencil tap test, and task orientation are between .75 and .91. The lowest at .75 is Object Pattern Reasoning and the highest at .91 is Task Orientation. This shows that the internal consistency of these tasks is within acceptable range. It is not surprising that Task Orientation demonstrated the highest internal consistency. This is a rating scale completed by the evaluator and often perceptions of development reported by evaluators or observers tend to demonstrate high correlations between items or domains.

No evidence of validity was reported for the ZamCAT. Although some of the tasks would be shown to have validity because they are already established measures.

Use: This assessment was developed as a part of the larger collaboration between the Zambian Ministry of Education, the Examination Council of Zambia, UNICEF, the University of Zambia, and the Center on the Developing Child at Harvard University which launched the Zambian Early Childhood Development Project (ZECDP) in 2009. The ZECDP is an effort to measure the effects of an ongoing anti-malaria initiative on children's development in Zambia. The intention was to develop a tool that could 1) provide internationally comparable measures of child development across domains; 2) be sensitive to local culture and linguistic differences; 3) be adapted for other developing countries (Fink, Matafwali, Moucheraud, and Zuilkowski, 2010).

Strengths:

Terms of Reference

1. ZamCAT covers a broad range of developmental domains.
2. Several of the domains assessed are less-commonly evaluated on large scale measures such as executive functioning and socio-emotional development.
3. The ZamCAT demonstrates that child development measures can accompany standard population-based household surveys.

Limitations:

1. The ZamCAT is limited in age use to preschool-age children.
2. A validity evaluation is needed. With so many alterations to the published measures and the development of new measures an examination of the validity is warranted.
3. The ZamCAT has only been used with children in Zambia (although it was used across several local languages).

4.2 The Ages and Stages Questionnaires (ASQ-3; 2009) is an example of a single parent rating scale that provides a very broad assessment of children's learning and development.

Purpose: The ASQ is primarily used to screen for developmental delays, but it has been used in research, as well.

Age: 1 month to 66 months (5 ½ years).

Format and administration: ASQ-3 is a developmental screening system comprising 21 age specific questionnaires for children between ages 1 month and 5 ½ years. Each questionnaire is completed by parents and includes a short demographic section and then 30 questions about the child's development. The child development questions are divided into five domains. Parents respond using the options of 'yes', 'sometimes' 'not yet'. Questions are phrased at a reading level for 4th-5th US school grade, which is roughly equivalent to a reading age of 9-10 years.

Developmental domains covered: The ASQ reports on communication, gross motor, fine motor, problem solving, and personal-social domains.

Time required: The ASQ takes approximately 10 to 15 minutes for a parent to complete and 2-3 minutes for professionals to score.

Training and materials: Little training is required for paraprofessionals or office staff to score the questionnaires. A User's Guide and training materials are available. Questionnaires, forms, letters, and activity sheets in the user's guides can be reproduced as many times as needed by a single site. Questionnaires are available in English or Spanish.

Scoring: The ASQ-3 results in a score (out of 60) for each area (communication, gross motor, fine motor, problem solving and personal-social) and these are compared to cut-off points on the scoring sheet. Scores beneath the cut-off points indicate a need for further assessment; scores near the cut-off points call for discussion and monitoring; and scores above the cut-off suggest the child is on track developmentally.

Technical properties: The ASQ-3 was standardized on 15,138 children in the United States whose parents completed 18,232 questionnaires. Families were educationally and economically diverse, and their ethnicities roughly matched estimates from the 2007 U.S. Census. Sensitivity (proportion of positives for

Terms of Reference

developmental delay correctly identified) was .86 and specificity (proportion of negatives for developmental delay correctly identified) was .85 overall. Figures for sensitivity and specificity at key ages between 24-30 months are given below:

At 24 months: sensitivity 91.2%, specificity 71.9%

At 27 months: sensitivity 77.8%, specificity 86.4%

At 30 months: sensitivity 86.7%, specificity 93.3%

The ASQ has been validated using the Bayley Scales of Infant Development II (BSID-II) and found to have a sensitivity of 100% and specificity of 87% at 24 months for severely delayed status.

Use: The ASQ-3 has been translated and used in a number of settings (e.g., France, Norway, Finland, Spain, the Netherlands, Turkey, North America, South America, Asia, and Australia). It has been used in studies with the general pediatric population and with children at increased risk for disability. Parents report that they find the questionnaires easy and quick to complete and they have been found to complete the questionnaire with reasonable accuracy.

Strengths:

1. ASQ-3 covers a broad range of developmental domains.
2. ASQ-3 produces scores (out of 60) for each domain and an overall score, which may allow measurement of small changes longitudinally.
3. Its format allows flexibility in administration. For example, for parents who may have difficulties with literacy or with language barriers, another individual could go through the items with the parent at the time of the administration. This would be a useful way of increasing access.
4. The authors comment that an important difference between this and other screening tools is that it is designed to show what children can do, not just what they cannot do.
5. The ASQ reports acceptable sensitivity and specificity.
6. The ASQ has been used among children at high risk of developmental problems.
7. It is quick and easy to complete and to score.
8. The ASQ is cost efficient as a one-off purchase with questionnaires and other materials being photocopied as required.

Limitations:

1. It has only been standardized in the USA so there is a lack of standardized norms for other populations.
2. In only a few studies has its psychometric properties been examined in their own cultural setting after translation.

Terms of Reference

3. ASQ-3 covers a broad range of developmental domains, but does not include social-emotional development, thus issues such as relationships are less well covered. However, ASQ-Social-Emotional focuses solely on social and emotional development, and could be used in conjunction with ASQ-3.
4. It is not clear whether it is valid to combine the scores from age specific questionnaires into one overall score.
5. Some of the language used in ASQ is 'Americanized'. Parents' understanding of this needs to be assessed and it possibly needs to be adapted for use in other settings.

4.3 Early Development Instrument (EDI; 1998). The EDI is an example of a rating scale completed by teachers or parents that has been very widely used internationally.

Purpose: The EDI is used as a screening tool for Kindergarten readiness.

Age: 4 to 7 years

Format and administration: The Early Development Instrument (EDI) is an assessment tool that provides a standard measurement that can help assess where children are and what areas need to be addressed to ensure that children start kindergarten ready to learn. Teachers complete a 104-item questionnaire on each student, for which they check whether or not students have met specific developmental milestones across five domains.

Developmental domains covered: The EDI reports on physical well-being, social competence, emotional maturity, language and cognitive development, communication, and general knowledge.

Training and materials: Training is a necessary preliminary step to EDI implementation. A copy of the EDI Guide should be provided to each teacher respondent. In addition, a training/information session will ensure accurate, consistent interpretation of items, as well as inform respondents about the purpose of data collection, how results will be used, and the logistics of the data collection process. Respondents with some training in the early childhood area will likely require only minimal training on the use of the EDI. Questionnaires are available in English or French.

Scoring: Average scores are calculated for the 5 domains and 16 sub-domains. Scores are used to identify percentile ranks. Scores also allow for an estimate of the overall percentage of children vulnerable in school readiness. Scores are categorized as follows:

- **On track (Very Ready)** - The total group of children who score in the best 25% of the site's distribution.
- **On track (Ready)** - The total group of children who score between the 75th and 25th percentiles of the site's distribution.
- **Not on track (At risk)** - The total group of children who score between the lowest 10th and 25th percentile of the site's distribution.

Technical properties: Since 1999, EDI data have been collected for more than 300,000 children ages 4–5 years in Canada and several other countries. A subset of the database, consisting of data collected from 2000 and later, has been analyzed to establish normative values for the EDI domains. The subset comprises 116,860 kindergarten children.

Terms of Reference

The EDI has also been tested to ensure its reliability and validity psychometrically (Janus and Offord, 2007). Internal reliability and test-retest reliability is high for each domain (ranging from .82 to .96). However, parent-teacher correlations were low (ranging from .36 to .64). Concurrent, external, and predictive validity have also been reported, and there are a wide range of correlations depending upon the type of validity and specific comparison. For example, emotional maturity domain of the EDI has a correlation of only 0.11 with *PPVT* scores and 0.73 with the social-emotional subscale of *First Step*, a comprehensive screener used for identifying developmental delays in preschool children.

Use: The EDI has been translated and used in a number of settings (e.g., some regions in the United States, Canada, Australia, Chile, Egypt, England, Holland, New Zealand, and implementation in Jamaica, Kosovo, Moldova, and Mexico). It is to be completed by teachers in kindergarten classes after several months of observations. Primary uses are the following:

- Serves as a population-level measure for interpreting outcomes for groups of children.
- Yields results that could be used by communities to identify weak and strong sectors.
- Encourages communities to mobilize and make plans to improve children's outcomes.

Strengths:

1. EDI covers a broad range of developmental domains.
2. EDI is an effective tool to assist decision-makers at various levels with resource planning for children.
3. EDI maintains consistent core concepts but it is culturally adaptable to local communities.
4. EDI is malleable to various populations.
5. The EDI is a helpful tool for determining school readiness.
6. Repeating data collection over time using the EDI in the same communities or regions makes it feasible to assess change.

Limitations of EDI:

The technical adequacy of the EDI is unclear. Particularly concerning are the strong differences between teacher and parent ratings. Hopefully, there will be additional information available in the near future.

4.4 GOLD (Teaching Strategies) illustrates a performance assessment that is widely used for formative purposes, but which also has been used as a summative assessment.

Purpose: Teaching Strategies GOLD tracks children's efforts, achievements, and progress. It is designed to inform instruction and enhancing learning outcomes.

Age: Birth through Kindergarten

Format and Administration: Teachers rate children's skills, knowledge and behaviours along a progression of development and learning. Scoring for GOLD varies slightly on the objectives. Objectives 1 through 23 are scored on a 0 to 9 point scale and objectives 24 through 36 are scored on a 0 to 2 point

Terms of Reference

scale. The scale has descriptions or “indicator levels” at scores 2, 4, 6, and 8 that describe the developmental continuum. The scores without “indicator levels” are used to document that the skill may be emerging but not yet fully established. This assessment is available in English and Spanish.

Developmental Domains Covered: GOLD evaluates social-emotional, physical, language, cognitive, literacy, mathematics, science and technology, the arts, and English language acquisition (where appropriate)

Time Required: Teachers collect evidence (anecdotal notes, work samples, photographs, etc.) of children’s development over a period of between 4 and 12 weeks. Then, use this evidence to score the children on the rating scale. In one survey of kindergarten teachers, the teachers reported using 1.5-3.0 hours of documentation time per child over an 11 week period (Williford, Downer, & Hamre, 2013).

Training and Materials: Training for the assessment ranges from one day in-person training to several days. Two days of in-person training is typical. There is also an online training option. Teachers using the instrument are offered optional participation in an inter-rater reliability certification. Here teachers analyze portfolios, score the data, and those scores are then compared with those of Teaching Strategies GOLD developers, with an agreement goal of 80% or better agreement.

Technical properties: Norms were calculated using a nationally representative norm sample of 18,000 children from 50 states, Puerto Rico, and Washington, DC and spans across age cohorts. GOLD provides norm tables across all six areas of development (Teaching Strategies, 2013). Each norm table includes expected scores for children across 24 different 3-month age bands from 0-71 months. There are norms for fall, winter, and spring. A study by Kim and Smith (2010) with infants through children aged 2 years showed high internal consistency reliability with a coefficient of .95-.99. This study also showed moderately high reliability. Teaching Strategies (Teaching Strategies, 2013) reports on GOLD’s concurrent validity with a study looking at preschool children. First, teacher rating scales of children’s social functioning and their learning behaviours related to the GOLD scale scores ($r = .426-.541$). Second, the related GOLD sub-scale scores for preschool children correlated low to moderately to standardized test scores on appropriate standardized assessment ($r = .307-.522$). These standardized assessments included the Peabody Picture Vocabulary Test, Pre-Language Assessment Scales, Woodcock-Johnson III Tests of Achievement, The Pencil Tapping portion of the Preschool Self-Regulation Assessment, and the Head-Toes-Knees-Shoulders Task. In an evaluation of GOLD in kindergarten (Williford, Downer, & Hamre, 2013), all teachers began the reliability process, 50% completed the tasks and 28% achieved reliability certification for all domains. Additionally, this research showed that the GOLD assessment was similar in scores to direct assessments in mathematics ($r = .64$) and literacy ($r = .53$), but children’s language and cognition were not as similar to the direct assessments of language ($r = .36$) and self-regulation skills ($r = .27$ and $.31$). Further, this study concluded that GOLD appears to measure children from different backgrounds equitably.

Use: Teaching Strategies GOLD and its predecessor, The Developmental Continuum, has been used extensively in the US in early education classrooms. Recently, GOLD has been used more consistently as a kindergarten entry assessment. In addition, Teaching Strategies is in development of a first through third grade assessment tool.

Strengths:

1. GOLD covers a broad range of developmental domains.
2. GOLD is a teacher-administered tool based on the child’s performance in his or her natural learning environment.

Terms of Reference

3. This assessment will help inform instruction for children.
4. GOLD is used widely in the United States.
5. GOLD has evidence of effective use with children with disabilities and dual language learners and appears to measure children from different backgrounds equitably.
6. GOLD provides normative data.
7. GOLD covers a large age span (currently birth through kindergarten with up to third grade in development).

Limitations:

1. Ongoing support of use in the classroom is generally needed when implementing an observation-based assessment system of this magnitude.
2. This systematic approach to assessment in the classroom environment may be new to many teachers and can be cumbersome at first.
3. This type of assessment should not necessarily be used alone for high-stakes decisions on individuals or programs.
4. The GOLD is normed on US students only.
5. GOLD reports varying psychometric properties depending on the study and age group of the children.
6. GOLD may more accurately assess math and literacy skills than other skills on the instrument.

4.5 The Hong Kong Early Child Development Scale (HKECDS; 2012) provides an example of a broad direct assessment developed outside North America designed to give a holistic assessment of child development.

Purpose: The HKECDS is used to assess the holistic development of preschool children as well as incorporating current expectations of early child development in Hong Kong. It can be used to evaluate the efficacy of targeted interventions and broader child-related public policies in early child development in Hong Kong.

Age: Preschool (ages 3 to 6).

Format and administration: The HKECDS relies on direct assessment with preschool-age children by a trained examiner. It is a developmental scale such that older children achieve higher scores in each learning domain.

Developmental Domains Covered: The HKECDS examines the following 8 learning domains with 95 test items: personal, social and self-care (9 items), language development (14 items), pre-academic learning (29 items), cognitive development (10 items), gross motor (12 items) fine motor (10 items), physical fitness, health and safety (9 items), and self and society (10 items).

Terms of Reference

Time Required: The total battery takes 30 to 45-minutes for all 95 items. The original version consisted of 190 items and required two testing sessions of 30 to 45-minutes each.

Training and Materials: Assessors in the validation study were undergraduate and graduate students majoring in early childhood education. Prior to formal data collection, they were required to go through all test items, instructions and materials with the second author. In addition, they had to achieve an inter-rater reliability of about 90% agreement before starting formal data collection. For each item, assessors use standardized stimuli and follow standardized instructions, procedures, and scoring rules. Gross motor activities are conducted either inside or outside the room, depending upon the space in the room. Specialized training is required for future assessors who wish to administer the HKECDS.

Technical properties: The HKECDS reports reliability information using the Chronbach's Alpha. This coefficient assesses the reliability of a test by examining the internal consistency. The following Cronbach's Alpha coefficients are reported for the domains: Personal, Social and Self-care .63, Language development .80, Pre-academic learning .95, Cognitive Development .70, Gross Motor .78, Fine Motor .75, Physical Fitness, Health and Safety .61, Self and Society .64. However, there were also moderate inter-correlations among subscales that assess theoretically different constructs.

Use: This instrument is used to assess the holistic development of preschool children as well as incorporating current expectations of early child development in Hong Kong.

Strengths:

1. The HKECDS covers a broad range of developmental domains.
2. Results from the validation study indicate that the HKECDS is a psychometrically robust, culturally and contextually appropriate measure of holistic child development for children ages 3 through 6.
3. Items in the HKECDS tap culturally sensitive expectations in each domain (for example, the measure examines young children's development of finger coordination with chopsticks, which is a unique activity in the Chinese culture).
4. Domains that are represented are translatable to children in different countries.

Limitations:

1. While this tool is valuable for its cultural sensitivity in Hong Kong, it is not adaptive to alternative populations.
2. The validation sample was not a representative sample of children in Hong Kong.

4.6 International Performance Indicators in Primary Schools (iPIPS) was explicitly developed as a broad assessment for use in international studies.

Purpose: The iPIPS was developed based on the Performance Indicators in Primary Schools (PIPS). It is used to assess what children know and can do, and how that changes during the first year of school. It also includes Personal Social and Emotional Development (BSED), Behaviour and Physical Development. It

Terms of Reference

collects information about prior educational experiences from parents or guardians and information about school from teachers. The intention is to use iPIPS in an international study to examine how one country compares to another at the start of school and after one year of schooling. Collecting data at this early stage also provides information to the extent that differences in later international studies (e.g., PISA) can be explained by differences in the early years.

Age: Used in the first year of formal schooling; at ages 4-7.

Format and administration: The direct assessment part of iPIPS is administered one-on-one either using a computer adaptive test or using a booklet with an application on a smart phone. Teachers complete a questionnaire rating children's Personal Social and Emotional development. Parental information can be collected using paper based questionnaires or over the internet.

Developmental Domains Covered: iPIPS directly assesses early reading, phonological awareness, early mathematics and short term memory. To be more specific, it includes: name writing (hand writing), picture vocabulary, ideas about reading, concepts of print, phonics, phonological awareness, letter identification, reading and word attack skills, word recognition and decoding skills, comprehension, early math, ideas about math, size and location, counting ability, simple number problems, digit identification using single, double and triple digits, and shapes. Optional items include: short-term memory, behaviour and attitudes. Teacher completed rating scales collect data on 12 aspects of Personal, Social and Emotional development and also can provide 18 items on behaviour. A parent survey collects basic data on children's physical development (height, weight, fine and gross motor coordination).

Time required: Direct assessment (computer or booklet) takes approximately 20 minutes. Additional time is required for the parent questionnaire/interview and supplementary surveys.

Training and materials: Time is required to become familiar with the user guide to get to know the computer system for the PIPS.

Scoring: Reports are generated by each country for schools using iPIPS. For the PIPS system the data are available online together with software to allow teachers to use the data. It is reported that it takes 30 to 60 minutes to access and interpret reports for the PIPS.

Technical properties: The Technical Report for PIPS version on a CD ROM from 2001 provides information about the reliability and validity of the instrument. It demonstrates test-retest reliability on 29 students who were re-assessed was 0.98 for the instrument. The subtests ranged from .34 to .99. Others have reported good reliability with different populations (Godfrey & Galloway, 2004). Predictive validity of the PIPS is demonstrated through the correlations ranging from .48 to .66 on assessments given up to 6 years later. The PIPS baseline assessment has been standardized to have a mean of 50 and a standard deviation of 10. The assessment provides conversion charts that offer age-corrected standardized test scores. Reliability and predictive validity data on the PSED and Behavioural part have also been published (Merrill & Tymms, 2001).

Use: The PIPS has been used in Australia, Netherlands, Scotland, New Zealand, Abu Dhabi, Germany, and South Africa. In England, PIPS has been widely used as on-entry assessment. 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).

Strengths:

1. Teachers and students reported to enjoy the computer delivery.

Terms of Reference

2. Teachers report that the program is easy to use.
3. The PIPS has been used wide-scale in several countries for 20 years across several languages.

Limitations:

1. Technology and IT support may be necessary to use the computer-based version.
2. Psychometric properties were found only for PIPS, not iPIPS, though it may be reasonable to extrapolate as they are highly similar. Psychometric properties must be established for each country.
3. The vocabulary and phonological awareness scales have proved the most challenging in terms of generating equivalent versions for the different languages and cultures, and it has been noted that these two scales should not be used for international comparisons (Tymms, Merrell, Hawker, & Nicholson, 2014).

4.7 Brigance Screens III (2013) is an example of a broad screening test relies on direct assessment by an educator or other expert in addition to observation.

Purpose: The Brigance Screens help educators identify potential developmental delays and giftedness, reduce over-referrals with at-risk cut-offs, determine each child's specific strengths and needs, and assess school readiness.

Age: 0–35 months includes Screens for Infants, Toddlers, and 2-Year-Olds

3–5 years includes Screens for 3-, 4-, and 5-Year-Olds

K & 1 includes Screens for 5- and 6-Year-Olds

Format and administration: Brigance Screens are administered by an educator (teacher, data collector, etc.). Educators spend only 10 - 15 minutes with each child in order to assess the first three domains (physical development, language, and academic/cognitive). These data are paired with parent and teacher observation of self - help and social - emotional skills to provide a quick snapshot of a child's skill mastery.

Developmental Domains Covered: The Brigance Screens assess physical development, language, academic/cognitive, self-help, and social-emotional skills.

Time required: Teachers spend approximately 10 - 15 minutes with each child, and then parents and teachers complete the assessment through observation.

Training and materials: Free online training is available on the publisher's website. Brigance Screens require very few resources to implement. Educators need the Screen Manual, a Data Sheet, and, for very young children, Screen accessories. For those sites that wish to enter the data on the online system, internet access is required.

Terms of Reference

Technical properties: Earlier versions of the Brigance Screens have demonstrated acceptable reliability and validity (Hamilton, 2006; Glascoe, 2002). Brigance Screens III (2013) also reports acceptable reliability and validity. The standardization of the assessment was conducted on a sample of children that was nationally representative in the United States in terms of geographic, demographic, and socioeconomic characteristics. Reliability is reported within acceptable ranges. Specifically, internal consistency is reported as .90 or higher, inter-rater reliability at .80 or higher and test-retest results were stable when tested at multiple points in time. Construct validity is demonstrated by the domain score structure of the assessment validated by confirmatory factor analysis. Differential item functioning analysis was used to examine for bias of gender and race along with a review panel these two methods showed no biased items. Content validity was reported by researchers and educators that the items on the assessment test the important developmental and early academic skills. The Brigance Screens III is reported by the publisher to correlate with other achievement, intelligence, and language tests such as the Vineland II and Woodcock Johnson III. However, exact correlations were not reported. Lastly, the publisher reports that the assessment correctly identifies the children with true developmental delays or disabilities demonstrating accuracy for sensitivity.

Use: The Brigance Screens are used widely in the US mostly in educational settings.

Strengths:

1. The Brigance Screens cover a broad range of developmental domains.
2. This screening assessment can be administered quickly.
3. This assessment spans a wide age range.
4. The Brigance Screens includes parents and/or teachers in rating.

Limitations:

1. It is available in English, Spanish, Laotian*, Vietnamese*, Cambodian*, Tagalog* (*For K&1 Screen, kindergarten level only.)
2. This assessment is generally administered by educators.
3. It was standardized on US children only.

4.8 The Kindergarten Entrance Inventory for Connecticut (KEI) is illustrative of teacher ratings of a type widely used in the United States for children entering primary school that offers relatively broad coverage of early learning and development.

Purpose: Kindergarten Entry Assessment to measure children's preparedness for kindergarten. Gives a state-wide snapshot of the skills and behaviours students demonstrate.

Age: 5-6

Format and Administration: Based on teachers' observations at the beginning of the kindergarten year. Teachers assign ratings on 6 domains that are defined by 3-5 indicators each.

Terms of Reference

Developmental Domains Covered: The KEI assesses language skills, literacy skills, numeracy skills, physical/motor skills, creative/aesthetic skills, and personal/social skills.

Time Required: Administration of this assessment requires time to observe the students to get to know them well enough for the teacher to complete this rating scale.

Training and Materials: The rating scale is the only material needed. It does not appear that much training is required for the teacher to rate the children (but the consequences for reliability and validity are unknown).

Scoring: The teacher rates each indicator in the domains on a scale of 1 to 3. Students at a score of 1, “demonstrate emerging skills in the specified domain and require a large degree of instructional support.” Students at a score of 2, “inconsistently demonstrate the skills in the specified domain and require some instructional support.” Students at a score of 3 “consistently demonstrate the skills in the specified domain and require minimal instructional support.”

Technical properties: The validity of the KEI was evaluated by comparing the content to the state preschool framework and curriculum. This comparison was reviewed by teachers in preschool and kindergarten. This instrument demonstrates a relationship to later grade 3 reading proficiency as assessed by the standardized state test.

Uses: The KEI is used in one state in the US as a comprehensive evaluation of children entering kindergarten.

Strengths:

1. The KEI covers a broad range of developmental domains.
2. Materials and training are not required.

Limitations:

1. The KEI I used only small-scale in the U.S.
2. This assessment is only designed for children at age 5.
3. Reliability and validity are largely unknown.

4.9 Evaluation of Potential for Creativity (EPoC; 2011) provides an example of an assessment designed to specifically measure important dimensions missing from many broad assessments.

Purpose: This assessment is used to measure two main modes of creative thinking.

Age: Elementary-middle school students (grades K-6)

Format and administration: EPoC includes two forms (A and B) to assess progress (pre- and post-test). Each form consists of 8 subtests which cover two domains of expression (verbal and graphic) as well as two modes of thinking [divergent-exploratory (D-E) and convergent-integrative (C-I) thinking]. For instance, divergent-exploratory verbal-type tasks, children generate ideas in response to one stimulus or problem (e.g., A DE verbal domain task is to propose as many story endings to a story beginning as possible within 10 minutes). In C-I graphic-type tasks, children are asked to produce an integrated,

Terms of Reference

elaborate and finalized composition (e.g., A CI graphic domain task is to generate an original drawing which combines a set of heterogeneous elements presented on a photo within 15 minutes).

Developmental domains covered: The EPoC examines creativity, cognition, and problem-solving.

Time required: Ten minutes are required for each divergent-exploratory task, and 15 minutes for each convergent-integrative task. There is also an allotted time warm-up activity before verbal tasks. This is a total of nearly 2 hours to complete this assessment.

Training and materials: Electronic version requires a computer (with recording for verbal tasks, and drawing program and a mouse for graphic tasks). The paper and pencil version requires only paper and pencil for the graphic tasks (verbal tasks are completed orally). Judges who score the integrative tasks are trained on the criteria and benchmarks for scores on the 7-point Likert scale, and then scores are compared among the judges.

Scoring: For divergent-exploratory tasks scoring is based on the number of ideas generated or a count the number of verbal or graphic productions. To score the convergent-integrative tasks, a 7-point Likert scale (1=low, 7=high) is used by independent judges to rate each drawing or story. Concluding the tasks, four scores are computed: Divergent-Exploratory thinking in the Graphic domain (DG), Divergent-Exploratory thinking in the Verbal domain (DV), Convergent-Integrative thinking in the Graphic domain (IG), and Convergent-Integrative thinking in the Verbal domain (IV).

Technical properties: EPoC was developed and validated with a sample of French students. Test scores were reliable with inter-subset correlations ranging from .60 to .78, and external validity was reported to be satisfactory (Baptiste, 2011). In one study, 48 Chinese children from a primary school in Hong Kong were tested for creative potential using the EPoC (electronic and paper& pencil version). For the electronic version, the Cronbach's alpha for verbal divergent-exploratory, verbal convergent-integrative, graphic divergent-exploratory and graphic convergent-integrative dimensions were .92, .83, .51, and .41. For the paper & pencil version, Cronbach's alpha for the graphic divergent exploratory and convergent-integrative dimensions were .76 and .65. A second study consisted of four groups (Chinese children in Hong Kong (HK), Chinese children in Paris, French children in HK, and French children in Paris) of primary school students (total of 540 children in grades 1-6) used the electronic version. Inter-rater reliability of verbal convergent-integrative dimension for HK-Chinese group, HK-French, Paris-Chinese and Paris-French was reported as .99, .95, .95, and .92 respectively. Inter-rater reliability of the graphic convergent-integrative dimension for the HK-Chinese, HK-French, Paris- Chinese and Paris-French group was reported as .99, .71, .98, and .96 respectively.

Use: This assessment was developed in France and is now available in several languages including French, English, German, Turkish, and Arabic. This tool has been used as a monitoring tool to guide creativity development.

Strengths:

1. First tool among creativity assessments that combines an approach by domain of creative expression and by mode of thinking, instead of measuring a single component.
2. It offers a broader vision of creative potential in children.
3. Available in several languages.

Terms of Reference

Limitations:

1. A relatively new instrument.
2. Administration time is long for this instrument.
3. Need wider use to examine the reliability and validity of the instrument with a larger set of children.

4.10 Thinking Creatively in Action and Movement (TCAM) illustrates an assessment focused on imagination, creativity, and divergent thinking which are rarely measured in broad assessments.

Purpose: Designed to measure fluency, originality, and imagination in young children without having to use written/ verbal responses. It was developed based on 4 guidelines: 1) kinesthetic (not verbal) modality is the most appropriate for eliciting creativity, 2) preschool children require procedures for warm-up and motivation, 3) tasks for assessing creativity should be things pre-schoolers are familiar with, 4) the test should be easy to administer and score.

Age: Preschool- primary (ages 3-8)

Format and administration:

Consists of 4 activities:

- Activity 1: “How many ways?”- assesses fluency and originality in moving alternate ways across the floor
- Activity 2: “Can you move like?”- assesses imagination in moving like animals or a tree
- Activity 3: “What other ways?”- assesses fluency and originality in placing a paper cup in a waste basket
- Activity 4: “What might it be?”- assesses fluency and originality in generating alternate uses for a paper cup.

TCAM is administered individually. The examiner should record all responses (in movement, in words, or both) made by the child as completely and accurately as possible. Only one child should be in the activity room at a time and they should have enough space for movement. Before administering, warm-up activities should be done. Examiners are encouraged to participate with the child when instructions are given and during the introductory phase of each activity.

Scoring: Scoring guide is provided in the test manual. Activity 2 is scored for Imagination and the other 3 activities are scored for Fluency and Originality. Fluency scores are the number of relevant responses, and Originality scores range zero- three points for each response (they are based on comparing responses to the statistical frequencies of responses in the originality lists in the scoring guide). Imagination scores are based on a 5-point Likert scale ranging from “no movement” to “excellent; like the thing.”

Developmental Domains Covered: The TCAM assesses motor, creativity, and cognition.

Time required: Administration takes 10-30 minutes (however no time limit should be imposed, the examiner should keep record of the time used).

Terms of Reference

Training and materials: Materials that are needed are: paper cups, wastebasket, pencils, red and yellow tapes.

Technical properties: Norms are based on 1,896 children ranging from ages 3-8 from 11 states and Guam. Inter-rater reliability is reported as coefficients between .90 and .99. Test-retest reliability is reported as .84 for a sample of 20 three-five year olds for a 2 week interval, and between .78 and .89 for a sample of 30, seven to eight year old boys with learning disabilities with a 1-14 day interval. Internal consistency was reported as .79. Significant positive correlations between TCAM and other creativity characteristics are reported. For example: correlations between the TCAM and production of various types of humour, between fluency scores and the Multidimensional Stimulus Fluency Measures, between TCAM and a modified Piaget measures of divergent thinking are reported. Scores on TCAM are showed only a low correlation to measures of intelligence. The TCAM results were not related to gender, socio-economic status, or race.

Use: This tool is used as a teaching tool. Teachers are more aware about the benefits of using creative movement in preschool and early elementary grades after using these tests.

Strengths:

1. This tool demonstrates acceptable reliability and validity.
2. This tool is easy to use for teachers.
3. The TCAM did not appear to be bias towards race, gender, community status, language/ culture.
4. Can examine abilities in young children and in children who are excluded from other testing instruments because of verbal restraints.

Limitations:

1. The TCAM has not been re-normed since 1981.
2. The originality lists associated with the TCAM have not been updated.
3. May not provide enough information about a child to make informed decisions or comparisons.
4. The assessment is designed as a teaching tool.

4.11 The Preschool Learning Behaviors Scale (PLBS; McDermott, Green, Francis, & Stott, 2000) and Learning Behaviors Scale (LBS; McDermott, Green, Francis, and Stott, 1999) illustrate relatively broad assessment of approaches to learning including motivation and executive functions.

Purpose: These scales were developed to examine the behaviours associated with learning.

Age: PLBS: Preschool age, 3-5; LBS: School age, 5-17

Format and Administration: The PBLBS has 29 items each presenting a specific learning-related behaviour. The teacher indicates whether the behaviour *most often applies*, *sometimes applies*, or *doesn't apply*. The items are varied with positive and negative learning behaviours to reduce response sets. The item content between the two measures is very similar with the wording altered for PBLBS to reflect less formal learning

Terms of Reference

contexts. Teachers rate the student as accurately as possible and should rate all responses. Teachers should have seen the student in school for at least 6 school weeks or 30 days.

Developmental Domains Covered: This assessment has four subscales: Competence Motivation, Attitude Toward Learning, Attention/Persistence, Strategy/Flexibility. Content focuses on attentiveness, responses to novelty and correction, observed problem solving strategy, flexibility, reflectivity, initiative, self-direction and cooperative learning.

Time Required: This scale takes teachers approximately 5 to 10 minutes per child to complete.

Training and Materials: There is no specific training involved. Materials needed include the rating scale.

Scoring: The evaluator calculates raw scores and converts them to percentiles. Students who obtain scores at or above the 40th percentile are displaying learning behaviours at or above the average range.

Technical properties: A factor analyses yielded distinct and reliable dimensions of competence motivation, attention/persistence, and attitude toward learning from several studies for both the PLBS and LBS across countries. However, in the U.S. it appears that the LBS presents a four factor structure. A normative sample (N=100) was configured based on the U.S. Census for PBLs. The normative sample for LBS was conducted with 1500 US students from 5 to 17 years old and was based on the 1992 U.S. Census. The assessment showed acceptable test-retest reliability and inter-rater reliability. In addition, the PLBS demonstrated expected correlations with the Social Skills Rating Scale (Gresham & Elliott, 1990) for concurrent validity evidence.

Uses: These scales are used in the US to examine children's specific learning-related behaviours. The PLBS has been translated to Spanish and tested in Peru. There is cross-cultural construct validity of the LBS as a measure of differential learning behaviours observed in school-aged children in Trinidad and Tobago. The tested dimensions of learning behaviours were found to be generalizable across age, gender and ethnicity.

Strengths:

1. Several studies supporting the validity of the instrument.
2. Assesses domains that are often neglected such as attitudes toward learning and persistence.
3. Used with several populations in various countries including US, Peru, Trinidad, and Tobago.

Limitations:

1. This is completed by a teacher who has spent significant amount of time with the child. Not all children attend school at an early age.
2. Standardization on US students only.
3. Narrow in focus by examining only learning behaviours.

4.12 Teacher Rating Scales of Early Academic Competence (TRS-EAC; Reid, Diperna, Missall, & Volpe, 2014) provides an example of a rating scale that combines measures of broad measures skills well beyond the academic sphere with measures of approaches to learning.

Terms of Reference

Purpose: A strengths-based measure to screen a wide array of skills, behaviours, and attitudes that are indicative of school success for preschool-aged children.

Age: Preschool aged-children, 3-5

Format and Administration: Includes two broad scales named the Early Academic Skills (39 items) and Early Academic Enablers (49 items). Teachers rate each child's current skill level compared with children of the same age.

Developmental Domains Covered: Early Academic Skills: Early literacy, early language, early mathematics, and early thinking. Early Academic Enablers: engagement, motivation, self-regulation, motor, interpersonal, and emotional competence.

Time Required: Time to complete the rating scales is not reported. With a total of 88 items and an estimated 15-20 seconds per question one can estimate that the total time on this assessment is less than 30 minutes per child.

Training and Materials: Teachers completing the rating scales would do best with a firm understanding of child development and appropriate age-level skills and behaviours.

Scoring: Teachers rate children on a Likert scale ranging from 1 (significantly below age expectations) to 5 (significantly above age expectations).

Technical properties: This assessment was evaluated with 440 preschool children from 38-70 months and completed by their teachers (N= 60). Most children, 62 percent, were Caucasian, 25 percent were Hispanic, 6 percent were African American, 1 percent Asian, and 6 percent classified as "other." All children were from lower socioeconomic backgrounds. Factor analysis supported a five-factor solution for Early Academic Skills Scale (Creative Thinking, Critical Thinking Skills, Numeracy, Early Literacy, and Comprehension) and a five-factor solution for the Early Academic Enablers Scale (Approaches to Learning, Social and Emotional Competence, Fine Motor Skills, Gross Motor Skills, and Communication). Experienced preschool teachers evaluated the rating scale for appropriateness and importance as an examination of content validity. Content validity ratios demonstrated acceptable levels of validity for the items.

To examine concurrent validity within two weeks of the teachers' rating of the children research staff individually administered achievement measures. Measures used for this correlation were the Woodcock-Johnson Tests of Achievement, 3rd edition (WJ-III; Woodcock, McGrew, & Mather, 2001), The Test of Early Reading Ability, 3rd edition (TERA-3; Reid, Hresko, & Hammill, and the Test of Early Math Ability, 3rd edition (TEMA-3; Ginsburg & Baroody, 2003). The TRS-EAC scales were associated with these direct measures. Factor scores from the rating scales were correlated with WJ-III Literacy and Math raw composite scores, the TERA-3 raw composite scores, and the TEMA-3 raw composite scores. Additionally, TRS-EAC scales were moderately predictive of subsequent performance for mathematics when the fall teacher ratings were correlated with the spring TEMA-3 raw composite scores for a small subsample of participants. Reliability of the scales was examined through the internal consistency of factors. Cronbach's Alpha ranged from .67-.98.

Uses: The TRS-EAC is used to assess the early academic competence for at-risk preschool populations.

Strengths:

Terms of Reference

1. This is a comprehensive assessment covering several domains including those not often measured in comprehensive assessments such as approaches to learning: engagement and motivation.
2. This is an easy to administer teacher rating scale.
3. The TRS-EAC is a strength-based measure rather than deficit-based.

Limitations:

1. The person completing the scale must be knowledgeable about appropriate age expectations in all of the domains.
2. The results are reported by teachers of preschool children which can limit the studied population to those that attend a preschool program.
3. This instrument does not appear to be available in multiple languages at this time.
4. This is a new assessment not widely used yet and more research with a wider population is needed.
5. It has a fairly narrow assessment age range. Available for only preschool-aged children.

4.13 Assessment of Peer Relations

Purpose: Designed to improve the peer-related social competence of young children. This assessment can be specifically of value to all children experiencing problems in establishing and maintaining successful and productive relationships with peers. Although for this assessment, peer relations are assessed in their school setting, both family and community factors are included in evaluation and intervention.

Age: 3-5 years old.

Format and administration: The assessment consists of three components. In the first section, one learns the general nature of the child's observed peer interactions in conjunction with an assessment of processes that allow for effective peer interactions to occur. Summary statements provide a bridge between assessment and intervention and special considerations such as possible developmental issues. This section consists of a series of scales to be observed while watching the child play, as well as written summaries and notes to determine developmental levels of the child. Scales are weighted rarely, sometimes, often, and almost always. At the end of the first section, assessors are asked to design interventions for the child to improve peer interactions. The second section involves observations of three social tasks important to young children (peer group entry, conflict resolution, and maintaining play). These three social tasks are evaluated using a checklist of behaviours to be observed and rated and surveys (with same scales as stated above). The purpose of this step is to evaluate how children think about a particular problem during interactions with peers. Next, an assessment is made of the child's ability to recognize specific social tasks and consistently and effectively perform those tasks over time. This is evaluated using charts that list concerns with emotional regulation, social cognitive processes and higher-order processes during play, and the child's different responses to these conflicts. The observer is to note how the child gains entry into a group of peers, how they resolve conflicts, and how they maintain play. Finally, a special considerations summary report related to the social tasks is provided.

Terms of Reference

Developmental Domains Covered: Communication, Problem solving, Personal-Social, Relationships with other children

Time required: Not reported, but appears to be a lengthy assessment.

Training and materials: Guide includes templates for observation.

Scoring: This assessment does not result in a numerical score. Rather, it is meant to be used as a tool for creating a specific intervention program for each child that is evaluated. It is also for the purpose of determining possible developmental disabilities. A “special considerations summary report” related to each child’s social tasks is generated from the assessment.

Technical properties: No information found on the standardization of this evaluation tool.

Use: Used by educators and also for clinical use. It is meant for both administrators to think about complex factors that influence young children’s peer relations, and intervention methods on how children can be helped in why they may be expressing difficulties in peer relations.

Strengths:

1. This assessment is meant to bring a clinical understanding and educational understanding to developmental issues.
2. The Assessment of Peer Relations gives a strong qualitative perspective of specific developmental issues with individual children.
3. Different factors such as family and community are considered in this measure besides classroom behaviours with this assessment.

Limitations:

1. No information available on standardization for this instrument.
2. This test does not give overall scores that could be used in comparisons.

4.14 Child Behavior Scale (*Excluded by Peers Subscale*) offers an example of a teacher rating scale solely focused on children’s peer relationships.

Purpose: To identify children who experience exclusion by peers.

Age: 5 to 13 years old (most commonly), but appears that it could work with younger children.

Format and administration: Teachers rate students as 0=doesn’t apply, 1=applies sometimes, and 2=certainly applies on the following seven items:

1. Peers refuse to let this child play with them.
2. Not chosen as a playmate by peers.
3. Peers avoid this child.

Terms of Reference

4. Is excluded from peers' activities
5. Is ignored by peers.
6. Not much liked by other children
7. Ridiculed by peers

Developmental Domains Covered: This is an assessment of peer relationships only.

Time required: Administered in 1-2 minutes per child.

Training and materials: This is a teacher report and no training is required.

Scoring: Lower scores on this scale indicate more positive peer relations. To score the scale, sum the items and divide by the number of responses. Because most children are generally accepted by peers, receiving a rating of 1 or 2 on just one or two of these items may raise concern.

Technical properties: This scale has been found to be valid and reliable for children ages 5 to 13.

Use: Used to identify children who are experiencing exclusion by peers.

Strengths of Measure:

1. This scale is quick and simple to administer.
2. It is easy to score and an easy way to identify possible areas of concern.
3. Scale is focused on an important area of concern for young children.

Limitations:

1. The scale has very few items.
2. Narrow in scope.
3. It assesses only problems not positive aspects of peer relationships.

4.15 Parenting Stress Index provides an illustration of a parent rating scale that assesses children's social-emotional development and relationship with the parent.

Purpose: The Parenting Stress Index is designed to be a screening and diagnostic measure to identify stressful aspects of parent-child interactions.

Age: Used for children 3 months to 12 years.

Format and administration: The assessment consists of 101 items with optional 19-item life stress scale. The short form has 36 items within 3 subscales: parental distress, parent-child dysfunctional interaction, and difficult child.

Terms of Reference

Developmental Domains Covered: The full version includes 6 child subscales (adaptability, acceptability, distractibility/hyperactivity, demandingness, mood, reinforces parent) and 7 parent subscales (competence, social isolation, attachment, parent health, role restriction, depression, relationship with spouse). There are also optional total stress scores and life stress scores.

Time required: The completion time for this index is 30-minutes for original and 10-minutes for short form.

Training and materials: Parents are to complete the assessment and no training is required.

Scoring: Total scores are calculated for each subscale.

Technical properties: Normed on several different samples including 534 parents of children in paediatric practice in Virginia, 191 low-income mothers in paediatric primary care clinics, and 223 Spanish-speaking mothers in NYC. Reliability for parents ranges from .55 to .80 and for children from .62 to .70. Test-retest reliability after 1 year was .70 for parent (.71 after 3 weeks) and .55 for child (.82 after 3 weeks). Low scores on the parent section correlate with parents having little investment in parenting or dysfunction in parent-child system.

Use: Useful in prevention and intervention programs, assessment of child abuse risk, and forensic evaluation for child custody.

Strengths:

1. This is simple and relatively quick to complete with no training required.
2. It is available in multiple languages: English, Dutch, Korean, Chinese, Portuguese, French Canadian, Italian, French, Icelandic, Japanese, Polish, Serbian, Swedish, and Greek.
3. There is a short version available.

Limitations:

1. May be difficult to get accurate information from parents who are defensive or have dysfunctional relationships with children.
2. Assesses a specific area of concern.

4.16 The Student Teacher Relationship Scale (STRS) is an example of a teacher rating scale for the teacher-child relationship.

Purpose: This was designed to evaluate teachers' feelings and beliefs about individual student's actions toward them, based on teacher perceptions of the teacher-child relationship.

Age: Appropriate for preschool to grade 3.

Format and administration: Using a 5-point Likert-type scale that ranges from 1 = definitely does not apply to 5 = definitely applies, teachers rated how applicable each statement is to their current relationship with a particular child. Three subscales are included in the measure. The Conflict subscale taps the extent to which the teacher-child relationship is marked by antagonistic, disharmonious interactions (e.g., "This

Terms of Reference

child and I always seem to be struggling with each other”). The Closeness subscale is an index of the amount of warmth and open communication present in the relationship (e.g., “I share an affectionate, warm relationship with this child”). The overall quality of the relationship is determined by the amount of closeness and conflict (reflected) in the relationship. The Dependency subscale measures the degree to which a teacher perceives a particular student as overly dependent on him/her. High dependency scores suggest that the student reacts strongly to separation from the teacher, requests help when not needed, and consequently the teacher is concerned about the student’s overreliance. Higher scores indicate more positive, higher quality teacher–child relationships. The items are based on attachment theory and the Attachment Q-Set (Waters & Deane, 1985).

Developmental Domains Covered: The full version includes 3 subscales – conflict (12 items), closeness (11 items), and dependency (5 items). The short form comprises 15 items that measure 2 dimensions of teacher-child relationships: Closeness and Conflict.

Time required: Time to complete this is 5 to 10-minutes for full version and 2-minutes for short form.

Training and materials: Teachers are to complete the assessment and no training is required.

Scoring: Each item is scored from 1 to 5. High total scores suggest higher teacher-child relationship quality, and specifically, a relative lack of conflict, lower dependency, and higher closeness.

Technical properties: The STRS was normed on a sample of more than 1500 students (and 275 teachers) that matched the 1990 US census data in terms of race/ethnicity and also reflected a wide range of socioeconomic status. It has also been shown to be psychometrically reliable and valid. Test-retest correlations over a 4-week period were .88 for closeness, .92 for conflict, and .76 for dependency. Validity studies indicate that the STRS correlates in predictable ways with concurrent measures of academic skills and performance on standardized tests (Hamre & Pianta, 2001).

Use: Primarily used as a tool for assessing student-teacher relationships in the context of efforts to prevent or to intervene early in the course of development of adjustment problems in school. The STRS can also be used in educational assessment batteries to determine the extent to which relationship problems or strengths should be addressed in program planning, and it can be used as a tool for researching classroom processes.

Strengths:

1. The STRS has been widely used in studies with preschool and elementary school children. It is associated with children’s and teachers’ classroom behaviours and correlates with observational measures of quality of the teacher–child relationship (e.g., Birch & Ladd, 1997; Howes & Hamilton, 1992; Howes & Ritchie, 1999).
2. STRS scores correlate with Attachment Q-Set ratings of teachers and students such that higher STRS scores are associated with more secure relationships (Howes & Ritchie, 1999).
3. This scale can be used with a preschool to grade 3 age range.

Limitations:

Only teacher perceptions are relied upon and children’s perceptions are not considered.

4.17 Early Years Foundation Stage Profile (EYFSP)

Purpose: The EYFSP was developed to inform parents about their child's development against the early learning guidelines and the characteristics of their learning, to support a smooth transition to key stage 1 by informing the professional discussion between EYFSP and key stage 1 teachers, and to help year 1 teachers plan an effective, responsive and appropriate curriculum that will meet the needs of all children.

Age: This assessment offers a two-year-old "check" between the ages of two and three and the EYFS profile is completed by the end of the year in which the child reaches age five.

Format and administration: The EYFSP profile summarizes and describes children's attainment at the end of the EYFS. Practitioners' assessments are primarily based on observing a child's daily activities and events. The assessor notes the learning which a child demonstrates spontaneously, independently and consistently in a range of contexts. Accurate assessment takes into account the perspectives of the child, parents and other adults who have significant interactions with the child.

Developmental Domains Covered: The EYFSP assesses 17 early learning goals in six areas of learning. **Communication and language:** listening and attention, understanding, speaking; **Physical development:** moving and handling, health and self-care; **Personal, social and emotional development:** self-confidence and self-awareness, managing feelings and behaviour, making relationships; **Literacy:** reading, writing; **Mathematics:** numbers, shape, space and measures; **Understanding the world:** people and communities, the world, technology; **Expressive arts and design:** exploring and using media and materials, being imaginative.

The measure also examines the child's three learning characteristics: Playing and exploring- engagement; Active learning- motivation; Creating and thinking critically- thinking.

Time required: The profile is completed over time after observations of the child in an ongoing process.

Training and materials: The Local Authority is responsible for training and supporting the teachers/practitioners. They provide support and guidance for all teachers/practitioners in making accurate assessments of children's achievements and progress through a range of strategies grounded in observations over time. It is unclear how much training is required to administer the profile.

Scoring: First, the report includes the child's attainment in relation to the 17 ELG descriptors. These are scored on a nine point scale. The first three points of each scale describe a child who is still progressing towards the achievements described in the early learning goals. The next five points are from the early learning goals themselves. They are not necessarily in hierarchical order and a child may achieve a later point without achieving some of the earlier points. The final point in each scale describes a child who has achieved all points one through eight and has developed further and is consistently working beyond the level of the early learning goals. These scores are categorized into emerging (1-3), expected (4-7) and exceeding (8-9).

Second, a short narrative describing the child's three characteristics of effective learning is generated by the assessor.

Standardization and psychometrics: Teachers are moderated in their use of this instrument. There are moderators that visit schools to sample students. The moderator secures consistency and accuracy of judgments made by teachers and assures that the setting has achieved an acceptable level of accuracy and validity. The moderator does this by evaluating several profiles to establish if the practitioner has understood what constitutes an appropriate outcome and judgment.

Terms of Reference

Analysis of data from the EYFSP indicated that six scales provide reliable measures of underlying skills. The simplest factor to measure uniformly is the Literacy factor. The least clear factor is Physical Development. The different scales appear to tap quite similar things as demonstrated by high correlations among domains. However, this may reflect how teachers make generalizations about pupils across domains and has been documented in other similar assessments. It was also reported that the EYFSP correlated with other language measures and was predictive of later achievement (Snowling, Hulme, Bailey, Stothard, and Lindsay, 2010).

Use: The EYFSP is used to inform parents of children's progress, to inform instruction in school, and to report children's progress nationally in England.

Strengths:

1. The EYFSP is being redeveloped to become more quantitative than qualitative (this new measure is not yet available).
2. The EYFSP is very comprehensive as it examines all key learning domains.
3. Although results are not reported on the moderation of the instrument, a strong program of moderating the use of the instrument is in place.
4. Used widely in England.

Limitations:

1. The EYFSP is being redeveloped to be more quantitative which could potentially compromise some domains that are currently evaluated (i.e., creativity).
2. The moderation of the instrument is likely an expensive endeavour.
3. The EYFSP is not currently used outside of England.

5. Conclusions and Recommendations

55. The assessments available offer many choices for measuring children's physical, social, emotional, linguistic, and cognitive development with respect to age, mode of assessment, the source or respondent, and burdens on respondents. There are fewer choices for assessments of executive functions and for some cognitive measures in the areas of math and science. Very few options are available for assessing development in the arts and culture and for approaches to learning; this is primarily done through performance assessments including clinical interviews (conversations and storytelling would be included here). Measures we reviewed that addressed aspects of approaches to learning including the specific topics of curiosity, creativity, critical thinking, and problem solving. None of the assessments we reviewed measured self-esteem, self-efficacy, values and respect, or subjective states of wellbeing such as happiness. We did not identify any comprehensive assessments for young children that addressed these domains.

56. For those domains that are measured rarely or not at all by comprehensive assessments, specific assessments are sometimes available. Most often these are rating scales (except for executive functions) completed by adults. Specific measures of wellbeing identified include those that assess relationships with parents and peers, and engagement and participation in ECEC. We did not identify measures of general

happiness and satisfaction specific to young children, but these could be constructed. In our opinion, general wellbeing and measures of children's rights to engagement and participation in decision making would be most readily assessed through clinical interviews or time diaries (with the latter requiring inferences from activities about quality of life as indicated by the engagement of children's capacities).

57. Clearly, some assessments have stronger evidence of technical adequacy than others. Concerns with technical adequacy are greatest for performance assessments and ratings, particularly in the domains that are not well-covered by tests. The technical adequacy of performance assessments can be improved by standardization of assessment procedures and training of assessors. This has costs, of course.

58. Given the available assessments, the most efficient strategy to selection of instruments for an international study would appear to be choosing one very broad assessment to be supplemented by a small number of highly specific assessments in domains that often are neglected. However, it would be possible to construct a broad assessment that is carefully tailored based on judgments regarding the best choice in each domain as was done with ZCAT. To fully cover all of the domains of interest to OECD representatives, some instrument construction may be necessary. Instrument adaptations will be required for language and culture. Given the extent of these adaptations, international pilot-testing to evaluate performance is recommended before use in a full scale international study.

5.1 Age

59. As the validity and reliability of assessment, and children's abilities to actively contribute to the assessment increase with age, the quality of the information obtained will be improved by conducting the assessment at age 3 or later. Prior to age 3, the assessments are primarily reports by adults. One then must choose whether to assess children at a particular age or at an educational transition such as entry to preschool or primary school. As entry to preschool can be well before age 3, this suggests entry to primary school as possible assessment point. However, this seems to us somewhat artificial, as what constitutes primary school in one country constitutes preschool in another. For this reason, we would recommend a uniform age across countries, perhaps age 4. One could also consider age 5 if by this age universal or near universal participation in ECEC (or primary education) has been achieved in the relevant country or countries, so that they can be assessed outside the home (including ratings by teachers). A final consideration is whether policy makers and others want just a point in time measure or wish to know how children are developing over time during the pre-primary years; in the latter case it will be necessary to administer comparable assessments at more than one age.

5.2 Final thoughts on decision making for national and international assessments

60. As pointed out previously, what, how and when children are assessed depends on the purpose or purposes of the assessment, judgments about what is important, and budgets. It is also limited by what is currently available, and some aspects of LDWB will require investment in assessment development if these are to be assessed on a large scale at an affordable cost. Whatever approach is taken, it would be wise to invest in some development, adaptation, and piloting before large scale use. Given the limitations of existing instruments, the most feasible course in the near future may be to administer a broad measure that addresses the domains identified as most important. This could be a single existing measure or a composite of existing measures. If some aspects of the measure are more costly (in time as well as money) to employ then it might be possible to administer those only to samples of the population or subsamples of a larger sample (matrix sampling in which only some items are administered to each child and then these are aggregated is another possibility, but it limits usefulness for teachers and policy studies). Lastly, we remind the reader that this paper reviews types and exemplars, and not all the available assessments. When selecting specific assessments for specific purposes, policy makers can consult experts in the relevant country or countries as well as the existing compendia or early childhood assessments.

Terms of Reference

REFERENCES

- Atkins-Burnett, S. (2007). *Measuring children's progress from preschool through third grade* (No. 5687). Plainsboro, NJ: Mathematica Policy Research.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Barbot, B., Besançon, M. & Lubart, T. (2011). Assessing creativity in the classroom. *The Open Education Journal* 4, 58-66.
- Barnett, W. S., & Boyce, G. C. (1995). Effects of children with Down syndrome on parents' activities. *American journal of mental retardation: AJMR*, 100(2), 115-127.
- Benítez, I., & Padilla, J. L. (2014). Analysis of nonequivalent assessments across different linguistic groups using a mixed methods approach: Understanding the causes of differential item functioning by cognitive interviewing. *Journal of Mixed Methods Research*, 8(1), 52-68.
- Berry, D.J., Bridges, L.J., & Zaslow, M.J. (n.d.). Early childhood measures profiles. Washington, DC: Child Trends.
- Borg, W. R., Gall, M. D., & Gall, J. P. (1989). *Educational research: an introduction* (5th ed.). New York: Longman.
- Creswell, J. W. (2008). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. (3rd ed.). Saddle River, NJ: Pearson.
- Dunn, L. & Dunn, L. (2006). *Peabody Picture Vocabulary Test, Fourth Edition (PPVT -IV)*. Bloomington, MN: NCS Pearson.
- Dunphy, E. (2008). *Supporting early learning and development through formative assessment: a research paper*. Dublin: National Council for Curriculum and Assessment.
- Ercikan, K. (2002). Disentangling sources of differential item functioning in multilanguage assessments. *International Journal of Testing*, 2(3-4), 199-215.
- Ettling, D., J. T. Phiri, et al. (2006). *Child Development Assessment in Zambia: A study of developmental norms of Zambian children aged 0-72 months* Lusaka, Zambia, Ministry of Education, Republic of Zambia.

Terms of Reference

Fernald, L. C., Kariger, P., Engle, P., & Raikes, A. (2009). *Examining early child development in low-income countries*. Washington DC: The World Bank.

Fink, G., Matafwali, B., Moucheraud, C., & Zuilkowski, S. S. (2012). The Zambian Early Childhood Development Project 2010 Assessment Final Report. Cambridge: Harvard University.

Frongillo, E. A., Tofail, F., Hamadani, J. D., Warren, A. M., & Mehrin, S. F. (2014). Measures and indicators for assessing impact of interventions integrating nutrition, health, and early childhood development. *Annals of the New York Academy of Sciences*, 1308(1), 68-88.

Glascocoe, F.P. (2002). The brigance infant and toddler screen: Standardization and validation. *Journal of Developmental & Behavioral Pediatrics* 23, 145-150.

Godfrey, J. R., & Galloway, A. (2004). Assessing early literacy and numeracy skills among Indigenous children with the Performance Indicators in Primary Schools test. *Issues in Educational Research*, 14(2), 144-155.

Hamilton, S. (2006). Screening for developmental delay: Reliable, easy-to-use tools. *Journal of family practice* 55, 415.

Hofferth, S. L., & Sandberg, J. F. (2001). How American children spend their time. *Journal of Marriage and Family*, 63(2), 295-308. Kim, D. H., & Smith, J. D. (2010). Evaluation of two observational assessment systems for children's development and learning. *NHSA Dialog*, 13, 253-267.

Korkman, M., U. Kirk, et al. (1998). NEPSY: A developmental neuropsychological assessment. San Antonio, TX, The Psychological Corporation.

Lau, Sing, et al. (2013). Bicultural effects on the creative potential of Chinese and French children. *Creativity Research Journal* 25, 109-118.

Matafwali, B. (2010). The relationship between oral language and early literacy development: Case of Zambian languages and English. Ph.D. Dissertation in progress. Lusaka, University of Zambia.

Melton GB. Young children's rights. In: Tremblay RE, Boivin M, Peters RDeV, eds. *Encyclopedia on Early Childhood Development* [online]. Montreal, Quebec: Centre of Excellence for Early Childhood Development and Strategic Knowledge Cluster on Early Child Development; 2011:1-8. Available at: <http://www.child-encyclopedia.com/documents/MeltonANGxp1.pdf>.

Merrell, C., & Tymms, P. B. (2001). Inattention, hyperactivity and impulsiveness: their impact on academic achievement and progress. *British Journal of Educational Psychology*, 71(1), 43-56.

Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741.

Rossbach, H. G. (1988). Daily Routines of Young Children. Paper presented at the Annual Meeting of the American Educational Research Association.

Sheridan, S., & Pramling Samuelsson, I. (2001). Children's conceptions of participation and influence in pre-school: A perspective on pedagogical quality. *Contemporary Issues in Early Childhood*, 2(2), 169-194.

Terms of Reference

Slentz, K.L., Early, D., & McKenna, M. (2008). A guide to assessment in early childhood: Infancy to age 8. Washington State Office of Superintendent of Public Instruction.

Snow, C. E., & Van Hemel, S. B. (2008). *Early childhood assessment: Why, what, and how*. Washington, DC: National Academies Press.

Snowling, M. J., Hulme, C., Bailey, A. M., Stothard, S. E. & Lindsay, G. (2011). Better Communication Research Project: Language and Literacy Attainment of Pupils during Early Years and through KS2: Does Teacher Assessment at Five provide a Valid Measure of Children's Current and Future Educational Attainments? (DFE-RR172A). London: Department for Education. Available <http://dera.ioe.ac.uk/13689/1/DFE-RR172a.pdf>

Standards and Testing Agency (2013). 2014 Early Years Foundation Stage Profile Handbook. London: Department of Education, Standards and Testing Agency.

Stevens, P. A., & Dworkin, A. G. (Eds.). (2014). *The Palgrave Handbook of Race and Ethnic Inequalities in Education*. Palgrave Macmillan.

Teaching Strategies (2013). Teaching Strategies GOLD Assessment System: Technical Summary. Washington, DC: Author.

Tinajero, A.R., & Loizillon, A. (2012). Early childhood development and wellbeing. The review of care, education, and child development indicators in early childhood. Paris: OECD.

William, F., & Monge, P. (2001). *Reasoning with statistics: How to read quantitative research*. (5th ed.). Belmont, CA: Thomson Higher Education.

Woodhead, M., & Brooker, L. (2008). A sense of belonging. Early Childhood Matters No. 111, 3-17. The Hague, The Netherlands: Bernard van Leer Foundation

Terms of Reference

ANNEX – MATRICES OF ASSESSMENTS

Matrix A – Assessment Descriptives

	Assessment - Author and Publication Date	Ages	Type of Assessment	Language	Countries Used in	Who administers
Assessor-report	Zambian Child Assessment Test (ZCAT) - Fink, Matafwali, Moucheraud, & Zuilkowski, 2010	Preschool	1-1 test	Nyanja, Bemba, Tonga, Lozi, English	Zambia; various subtests used in multiple countries	Trained assessors
	Battelle Developmental Inventory (BDI) - J. Newborg, J.R. Stock, J. Wnek, J. Guidubaldi, and J.S. Svinicki, 1988	Birth to 7 years 11 months	1-1 test; also parent and teacher interview items			Parents or Examiners
	NIH Measures - NIH Blue print of Neuroscience Research. Principal Investigator: Dr. Richard Gershon, 2004	3-85 years	Proctored; self-administered; computer-administered	English and Spanish	US , Colombia	Trained assessors
	Brigance Early Childhood Screens (BECS) - Albert H. Brigance, 1999	Birth to 68 months	Child observation and performance; also parent interview items		US	Professionals with child development knowledge
	Denver II - 1990	Birth to 6 years	Direct observation; also parent observation	English and Spanish		Professionals or para-professionals
	Griffiths Mental Development Scales Extended revised (GMDSER) - 2006	2 - 8 years	1-1 test	English	UK	Pediatricians and health professionals

Terms of Reference

	Assessment - Author and Publication Date	Ages	Type of Assessment	Language	Countries Used in	Who administers
	Mullen Scales of Early Learning (MSEL)- 1995	Birth to 68 months	1-1 test		US	“highly trained” professionals
	Schedule of Growing Skills (SGS) - 1996	Birth to 5 years	1-1 test		UK, England	Trained assessors
	Hong Kong Early Childhood Development Scale (HKECDL) - Nirmala Rao, Sun Jin, Sharon Ng, Kitty Ma, YvonneBecher, Diana Lee, Carrie Lau, Dr. CB Chow, & Patrick Oppen (1992, 1996)	3-6 years	1-1 test	English, Cantonese, Chinese	China (Hong Kong)	Trained assessors
	Early Years Foundation Stage Profile (EYFSP) – Snowling, Hulme, Bailey, Stothard, & Lindsay, 2011	2-5 years	Observation based assessment	English	England	Trained Teachers
	Early Learning System (ELS) - Riley-Ayers, Stevenson-Garcia, Frede, and Brenneman 2012; Riley-Ayers, Stevenson-Garcia, Brenneman, Thompson, & Thompson, 2014; Developed at NIEER	3-6 years	Authentic observation-based assessment	English	US, China	Teachers
	International Performance Indicators in Primary Schools (iPIS) - Peter Tymms and Colleagues: http://www.ipis.org/the-team Note: could be listed as direct assessment for cognitive domains.	4-7 years (first year of school)	1-1 test with teacher rating and supplemental parent report	Dutch, German, Russian, Spanish, French, Slovenian, Chinese, Afrikaans, Sepedi	Australia, Netherlands, Scotland, New Zealand, Abu Dhabi, Germany, South Africa	Trained teachers
	Work Sampling System (WSS) - Meisels, Jablon, Dichtelmiller, Dorfman, & Marsden, 1998	3yrs to sixth grade	Checklists used 1-1 or in group setting	English	US	Teachers

Terms of Reference

	Assessment - Author and Publication Date	Ages	Type of Assessment	Language	Countries Used in	Who administers
	Teaching Strategies GOLD - Teaching Strategies, 2010	Birth through Kindergarten	Authentic observation-based assessment	English, Spanish	US; Other countries, but do not have details about which or how many---have reached out to a contact at GOLD to inquire.	Teachers
	High Scope Child Observation Record (CORE)- High Scope, 2013	Birth through Kindergarten	Authentic observation-based assessment	English, Spanish	US, Canada, Chile, Indonesia, Ireland, Korea, Mexico, The Netherlands, Portugal, South Africa, UK	Teachers
	Early Development Instrument (EDI) - 1998	4-7 years	Questionnaire done by teachers or parents.	English, French	Canada , US, Australia, Chile, England, Holland, Egypt, Mexico, Jamaica,	Teachers, Early childhood educators
	Kindergarten Entrance Inventory for Connecticut (KEI – Connecticut)	K	Observation-based assessment	English	US	Teachers
Parent-report	Child Development Inventory (CDI) - Harold Ireton, 1992	15 months to 6 years	Parent report with professional assistance	English and French	US, France, Canada	Parents

Terms of Reference

	Assessment - Author and Publication Date	Ages	Type of Assessment	Language	Countries Used in	Who administers
	Ages & Stages Questionnaire (ASQ) - Bricker, D., and Squire, J., 1999	1 month to 66 mos.	Parent report	English, Spanish, French, Korean	France, Norway, Finland, Spain, Netherlands, Turkey, North America, South America, Asia, Australia	Parents
	Parents' Evaluation of Developmental Status (PEDS) - 1997	0 to 8 years old	Parent report	English, Spanish, Vietnamese, Hmong, Somali, Chinese, Malaysian	US, Australia, Great Britain, England	Parents

Terms of Reference

Matrix B – Assessment Details

	Assessment	Administration time	Cost per administration	Assessment format (electronic/paper & pencil)	Strengths	Weaknesses
Assessor-report	ZCAT	1 hour	?	Paper and pencil	Comprehensive, culturally specific	Evidence of validity limited to original assessments; long administration
	BDI	45 - 90 minutes	\$312.50	1. Direct assessment with toys, games, tasks 2. Observation 3. Parent report	It covers all the developmental domains, and adaptations are available for children with disabilities.	Training is required to administer, measure takes 1.5 hours, not standardized for use in the UK, caution is required for using with kids not from US, does not involve parents, evidence is lacking about acceptability by parents, relatively costly.
	NIH	9 tests + 2 supplemental tests. Time ranges from 2-7 mins for each test.	No cost for assessment. Fees apply for user & tech support: >100 subjects = \$1500 or <100 subjects = \$5000 covered until 12/31/2014 if non NIH support. If Study is NIH supported then the fees are paid as long as it is funded by NIH.	computerized	Easily incorporates multiple areas of neurological function; Inexpensive, no royalties, low per subject costs.	
	BECS	10-15 mins	\$149	paper/pencil; computer scoring available	Covers the developmental domains of interest, quick and easy to use, has acceptable psychometric properties, flexibility with administration.	More focused on academic performance, used in educational settings and not in health settings, has not been standardized in the UK; primarily used as screener
	Denver II	20-30 mins	\$90 for English packet, \$120 for Spanish packet		Covers the developmental domains, is well known and widely used, is reported to have good sensitivity.	Training is required, meticulous administration, has poor specificity, the measurement was standardized in 1980 so the norms are outdated and specific to US; used primarily as developmental screener.
	GMDSER	50-60 mins		Kit: building blocks, drawing book, record forms	Covers the developmental domains but personal-social rather than social-emotional, widely used, acceptable sensitivity and specificity, standardized recently on a UK population.	Intensive training required, no evidence for use as a population measure, no evidence about acceptability by parents, lengthy to administer, little published evidence on validity.

Terms of Reference

	Assessment	Administration time	Cost per administration	Assessment format (electronic/paper & pencil)	Strengths	Weaknesses
	MSEL	+/- 30 mins		Electronic and paper/ pencil	Covers the developmental domains and is relatively easy to score.	Professionals need to be highly trained, not a population measure, standardized 30 years ago, excludes children with disabilities, no UK norms, not used in general child developmental assessment, unknown how parents feel about it, parents are not involved, costly.
	SGS				Covers the developmental domains, widely used, the original specificity is good, completion time is relatively short.	The original estimates of sensitivity range from poor to good depending on the domain, the validity was carried out over 30 years ago so is now outdated, although SGS is being used in the Flying Start program, there is no published evidence of this, does not involve parents, no information available on acceptability by parents or professionals, little use in reviewed journals
	HKECDS	2 sessions of 30-45 mins	No info	Direct assessment: verbal and physical actions.	Considers both holistic development and current expectation of early child development in Hong Kong.	No representative sample of Prek school children in HK; Recruited different school children from different PreKs and family background of 2 pilots' sample size were small.
Teacher-report	EYFSP	Ongoing over the course of the school year		Observational; paper/pencil	Currently being redeveloped to become more quantitative; comprehensive in examination of all key learning domains; used widely in England	Moderation of the instrument is expensive; not currently used outside of England.
	ELS	Ongoing; Generally with 3 score periods per school year.	\$250.00 for print kit for 25 children; Online price dependent on the number of children assessed: range as high as 21.95 plus one time set up fee (\$450-\$900)	Paper and Pencil and Online	Measure that examines child development across several domains through a manageable assessment system for teachers. Provides valuable information for teachers to inform instruction. Provides a developmental trajectory of children's development.	ELS does not examine development of children in the arts; It is not used wide-scale at this time.
	iPIPS (only part is teacher report)	20-minutes for direct assessment (cognitive) portion	In England PIPS is £80 per school plus £3.10 per pupils	computerized for the objective part and booklet with app	Teachers and students reported to enjoy the computer delivery; program is easy to use; has been used wide-scale in several countries and in several languages for 20 years.	Technology and IT support may be necessary for some users; psychometric properties extrapolated from PIPS to iPIPS; vocabulary and phonological awareness scales have proved most challenging in terms of generating equivalent versions for the different languages and cultures.
	WSS	Ongoing	\$75 for basic Teacher Reference Pack and goes up from there	Paper and pencil	Comprehensive; charts children's progress overtime; gives insight for individualizing instruction	WSS may vary depending on teacher experience and training; sample used for reliability and validity was not representative of US population

Terms of Reference

	Assessment	Administration time	Cost per administration	Assessment format (electronic/paper & pencil)	Strengths	Weaknesses
	GOLD	Ongoing; Generally with 3 score periods per school year.	\$199.00 for print kit for 25 children; Online price dependent on the number of children assessed: range as high as 21.95 plus one time set up fee (\$450-\$900)	Paper and Pencil and Online	Comprehensive; developmental trajectory of learning and development; gives insight for individualizing instruction	May vary depending of teacher experience and training (teacher bias evident in one study); Large number of items to collect data on and evaluate; Minimal concurrent/ construct validity with direct assessments of language, cognition, and self-regulation.
	CORE	Ongoing; Generally three score periods per school year	\$225 for kit for 24 children paper and pencil administration	Paper and pencil and Online	Comprehensive; developmental trajectory of learning and development; gives insight for individualizing instruction	May vary depending on teacher experience and training; no normative data
	EDI	15 mins per child. Administered by teacher in 2 nd half of the year.	No info	Electronic (e-EDI) or paper/pencil	Emphasizes community aspects of readiness; can be adapted for different locations; considers both teacher and parent observation; predicts reading and writing outcomes in later years	Limited use in the U.S.; narrow age applications; cultural modifications need to be arranged with a professional assessor; limited agreement with other instruments with similar purpose; results reported at population level only, not for individual children
	KEI CONNECTICUT	Ongoing; generally observations at beginning and end of the kindergarten year.		Paper and pencil	Evaluates children comprehensively across several domains; minimal training and materials required.	Only used small-scale in the US, only appropriate for children entering kindergarten.
Parent-report	ASQ	10-15 mins	\$190	Paper and pencil/ electronic	Covers developmental domains and produces scores for each, can be a population measure, flexibility to administer, gets parents involved, provides a good basis for discussion, can be used with children at high risk of developmental problems, quick and easy, one-off purchase	Lack of standardized norms for the UK population, focuses on personal-social instead of social-emotional, lack of info about UK parents accepting the measure, need further analysis to see if it can be used with parents with language barriers and literacy problems, age specific questionnaires, and some of the language is "Americanised"; used primarily as screener
	PEDS	5 mins and 2 mins to score	\$60 for 100 survey forms	Interview or paper and pencil or electronic	Covers the developmental domains, used in population surveys, encourages parent involvement, flexible format, quick and easy, requires minimal training, low cost, and can be used among children at risk of developmental problems	May be hard to track small changes overtime, hard to discriminate in between children, less useful as a population monitoring measure, may be subjective, it assumes parents' knowledge about development; used primarily as screener.
	CDI	30 - 50 mins to complete and 10 mins to score	Manual \$30; Booklet \$10-\$15; Answer sheets \$10; Profiles \$10	Paper and pencil	Covers the developmental domains of interest, low costs involved, parents find it easy to complete, has been used with high-risk kids.	It is not a assessment to find developmental delay, but for kids who may have one, the standardization was over 30 years ago and with a largely white US sample, there are no UK norms for this measure, authors state it is not appropriate for parents with less than a high school degree; used primarily as screener.

Terms of Reference

Matrix C – Assessment Psychometrics

	Assessment	Purpose of Assessment (summative/formative)	Purpose of assessment (clinical/screening)	Normative data	Reliability	Validity	Use of Scores
Assessor-report	ZCAT	Summative	Research tool	Administered 1686 Zambian preschool children in 2010	Cronbach's alpha	Validity for individual assessments but not for full adapted battery	Census
	BDI	Summative	Depicts child progress in intervention programs; identifies children with special needs; provides comprehensive analysis of functional capabilities.	Normative data were gathered from 2,500 children (closely resembling the 2000 US census) btwn the ages of birth to 7yrs 11mos	Test-retest $r = .71-1.0$	Moderate correlations with other established tests; concurrent validity .566 with PPVT-R; .66 with Preschool Language Scale	Comprehensive test used by professionals to assess the development of children. Has been used among children with autism, developmental delays, motor delays, speech and language delays and prematurity.
	NIH	Summative	Research/clinical/longitudinal/epidemiological	Administered to 4,859 participants 3-85 y.o. Normative data available for ages 3-17, 19-29,30-39,40-49,50-59,60-69, 70-79, 80-85.	Strong	16,000 subjects for response theory approach. 450-500 subjects compared with Gold standard measures where available.	Used for variety of settings by researchers and clinicians with emphasis on measuring outcomes for longitudinal, epidemiologic studies + prevention and intervention trials.
	BECS	Summative	Screening to identify weaknesses, assess readiness for school; identify interventions needed	Sensitivity 82%, specificity 84%. Also identifies 86% of children under the age of 2 with potential giftedness.		Comparison with a battery of age-appropriate developmental assessment tools such as Bayley's.	Identifies potential learning delays or academic giftedness

Terms of Reference

	Assessment	Purpose of Assessment (summative/formative)	Purpose of assessment (clinical/screening)	Normative data	Reliability	Validity	Use of Scores
	Denver II	Summative		The authors present norms based on the 1980 US census, which is criticized.		Authors made no attempt to measure the validity of the tool.	To detect developmental delay. The Denver II is widely used in clinical settings and as the gold standard which other measures are compared.
	GMDSER	Summative	Clinical and research purposes	The measure was normed on a national representative sample of children in UK.			To measure the rate of development of young children, for clinical and research purposes.
	MSEL	Summative		Normative sample as based off of a US sample of 1849 children over 8 years.		Little evidence found for validity.	Now commonly used as a measure of cognitive language skills. Used in research, clinical evaluations, and longitudinal evaluation of children with autism.
	SGS	Summative				Validity of the original tool was assessed by comparison with the Griffiths test with the NCES tool.	To establish children's developmental level: widely used to assess children as they enter the program Welsh Flying Start and when they leave at 3 years of age.
	HKECDS	Summative	Developmental readiness.	Sample in Hong Kong is largely comparable with the Canadian normative references in the physical, social and emotional domains, but higher in the language/cognitive and communication/general knowledge domains.	The test-retest reliability of the CEDI after a four-week interval was analyzed in 30 participants using the kappa statistic (k). The kappa coefficient was 0.89, thus demonstrating the instrument's stability over time.		
Teacher-report	EYFSP	Formative	To inform parents about their child's development against the early learning guidelines and the characteristics of their learning.	Teachers are moderated in their use of the instrument. Moderators visit schools to sample students and evaluate several profiles to establish whether practitioners understand what constitutes an appropriate outcome and judgment.	Six scales have proven to be reliable measures of underlying skills. The simplest factor to measure is literacy and the least clear to measure is physical development.	The EYFSP has been correlated with other language measures and is predictive of later achievement.	Scores are used to inform parents of child progress, inform school instruction, and report child progress nationally in England.
	ELS	Formative	Track child's efforts, achievements, and progress; designed to enhance instruction and improve learning	This assessment is not normed.	High inter-rater reliability.	Moderate correlations with other established tests. High internal consistency.	Inform instruction and track children's progress over time

Terms of Reference

	Assessment	Purpose of Assessment (summative/formative)	Purpose of assessment (clinical/screening)	Normative data	Reliability	Validity	Use of Scores
	iPIPS (only part is teacher report)	Formative and Summative	Assess children's abilities and development at entry to school and assess progress during the first year of school.	The PIP baseline assessment has been standardized to have a mean of 50 and a standard deviation of 10. The assessment provides conversion charts that offers age corrected standardized test scores.	Test-retest reliability for 29 students was .98 for the full instrument; subtests ranged from .34 to .99.	Predictive validity has been demonstrated through correlations ranging from .48 to .66 on assessments given up to 6 years later.	Used to assess children at the start of school and after one year of schooling within and across countries.
	WSS	Formative	Track child's efforts, achievements, and progress; designed to enhance instruction and improve learning	Scores provided for diverse, but not standardized sample. Norms based off of 5 public schools in Pittsburgh over 3 years	Internal consistencies range from .87 to .94; high inter-rater reliability	High correlation between WSS and WJ-R subscales	Inform instruction and track children's progress over time
	GOLD	Formative	Track child's efforts, achievements, and progress; designed to enhance instruction and improve learning	Nationally representative norm sample of 18,000 children from 50 states, PR, and DC; across age cohorts; Provides norm tables across all six areas of development. Each norm table includes expected scores for children across 24 different 3-month age bands from 0-71 months. Includes norms for fall, winter, and spring.	Strong internal consistency; Inter-rater reliability for kindergarten teachers seems weak;	Varying validity based on study report and age of children; seems to be most valid for math and literacy; low concurrent and construct validity with direct measures of language, self regulation, and cognition.	Inform instruction and track children's progress over time; Statewide progress assessed at kindergarten entry in US
	CORE	Formative	Track child's efforts, achievements, and progress; designed to enhance instruction and improve learning	None?	85.7% agreement for inter-rater reliability for 70 teachers; acceptable internal consistency	Demonstrated significant differences in scores at age category and moderate to high correlations with other standardized measures.	Inform instruction and track children's progress over time
	EDI	Summative	Assesses children's level of development during the 1 st year of school and readiness to learn.	116,860 4-5 year old children from Canada + several countries. Data available @Offord Centre website.	Internal: .84-.96; Inter-rater: High between teachers and early childhood professionals, but moderate between these and parents.	International comparison of normative data is valid.	To determine the % of children at various levels of readiness to benefit from school; also plan interventions and resource investments

Terms of Reference

	Assessment	Purpose of Assessment (summative/formative)	Purpose of assessment (clinical/screening)	Normative data	Reliability	Validity	Use of Scores
	KEI - CONNECTICUT	Formative	Gives a statewide snapshot of the skills and behaviors students demonstrate			Validity evaluated by comparing the content to state preschool framework and curriculum was reviewed by teachers in preschool and kindergarten. Demonstrates relationship to later grade 3 reading proficiency.	Gives comprehensive evaluation of children entering kindergarten.
Parent-report	ASQ	Summative	Comprehensive screening program, used to identify children who need additional evaluations	Standardized on 15,138 children from families of different educational and economic backgrounds	Inter-observer agreement $r=.92$; Test-retest $r=.95$	The ASQ was validated against the Bayley Scales of Infant Development. Concurrent $=.84$; Sensitivity $=.72$; Specificity $=.86$	Screening for developmental delay
	PEDS	Summative	A surveillance tool and screening test to elicit parents' concerns about their child's development and health	It was standardized on 2823 families in the US from various backgrounds		PEDS has been compared with 14 other developmental assessment including Bayley's.	To elicit parents' concerns about their child's development and health.
	CDI	Summative	Screening requiring in-depth developmental information from parents.	The CDI came from 30 years of research with the Minnesota Child Development Inventory		Concurrent validity with IQ and achievement.	To screen and assess children where there are concerns about development, or follow high-risk children

APPENDIX TWO - IMPROVING EARLY LEARNING OUTCOMES – THE CONTRIBUTION OF INTERNATIONAL COMPARATIVE DATA

TABLE OF CONTENTS TOC \O "1-3" \T "SUMMARY HEADING,1,PREFACE HEADING,1,CONCLUSION HEADING,1,BIBLIOGRAPHY HEADING,1,APPENDIX HEADING,1,ANNEX HEADING,1,GLOSSARY HEADING,1,INDEX HEADING,1,EXECUTIVE SUMMARY HEADING,1,FOREWORD HEADING,1,ANNEXES HEADING,1,PART HEADING,1,INTRODUCTION HEADING,1" \H

Executive summary	89
Introduction	90
Trends in ECEC	91
Participation rates and investment in ECEC are continuing to increase	91
Improvements are being pursued in the quality of staffing	91
Curricula are also being redesigned	92
And monitoring and data collection is increasing	92
What type of ECEC provision can really make a difference?	92
ECEC can have enduring, positive effects	92
The impacts are greatest for disadvantaged children	94
But the benefits of ECEC may be reducing over time	94
Evidence on the factors that support good learning outcomes is weak	95
Are there significant data gaps?	95
A conceptual framework on early learning outcomes	97
Child outcomes were included in the OECD's programme of work and budget for 2015/16	97
A conceptual model of the relationship between ECEC and child outcomes	98
How would international comparative data add value?	103
For policy makers and policy advisors	103
For ECEC practitioners	104
For Parents	104
For other stakeholders	104
Developing a proposal	105
Decision to enter a scoping phase	105
Guiding principles	105
HYPERLINK \l "_Toc435530686" Scoping the study	106
HYPERLINK \l "_Toc435530687" Risk management	106
Approaches already in place	107
Early learning assessment for developing countries	108
Links to other OECD assessments and surveys	108
The ECEC staff survey	109
Education and Social Progress	109
Education for 2030	109
PISA	109
Indicative costs	110
Proposed timeline and next steps	111
REFERENCES	112

Terms of Reference

IMPROVING EARLY LEARNING OUTCOMES:
THE CONTRIBUTION OF INTERNATIONAL COMPARATIVE DATA

Executive summary

61. High quality early childhood education and care (ECEC) offers immense opportunities for children to develop the cognitive and social and emotional skills needed for their development, well-being and success in later life. High quality ECEC also offers countries a means to lift the performance of their overall education systems, and to mitigate disadvantages experienced by particular groups of children.

62. Increasing participation in ECEC, however, does not guarantee improved learning outcomes for children and young people. Some ECEC provision has long-lasting, positive benefits and some does not. And if early learning is impeded, it is difficult to compensate for this at a later stage.

63. ECEC sectors in many countries have experienced sustained change over the last two decades. Participation rates have risen, resulting in new and expanded services and higher numbers of ECEC staff. Countries are increasingly emphasising the educational components of ECEC, through improvements to staff qualifications, curricula and quality assurance processes.

64. The empirical data available on ECEC largely compares the impacts of participating in ECEC with not participating in ECEC, or the impacts of individual programmes on child outcomes. There is an absence of empirical data on the type and nature of ECEC participation and provision that best enhance early learning outcomes, including for different groups of children.

65. Gaining reliable information on what is currently being achieved and what could be achieved from ECEC will support policy makers, ECEC practitioners and parents in making informed decisions and in improving the value gained from ECEC for children.

66. To address the current gaps in information, the Programme of Work and Budget for 2015/16 included a new project on early learning outcomes. The European Commission has expressed its interest in collaborating with the OECD on this work.

67. The project will draw on the world's best available expertise in this area and will be scoped in close collaboration with participating countries, to ensure it is shaped by countries' policy objectives. These countries may choose to align the measures of early learning with the Programme for International Student Assessment (PISA), to show the links between early learning outcomes and later student success.

68. The learning outcomes that would be the subject of an OECD programme in this area are likely to be broad, and include:

- Social and emotional skills
- Communication and language skills
- Other cognitive skills
- Physical development
- Health, hygiene and nutrition
- Well-being.

Introduction

69. The purpose of this paper is to assess the contribution that international comparative measures of early learning outcomes could make to improving ECEC policies and practices. Such information would help parents, educators and policy makers to better understand the level and types of ECEC participation that are most beneficial to different groups of children, and how ECEC provision and home learning environments can best complement each other.

10. The definition of ECEC used in this paper is consistent with previous OECD work on ECEC. It includes all arrangements providing education and care for children under compulsory school age, regardless of setting, funding, opening hours, or programme content (OECD, 2006).

11. The key types of ECEC covered by this definition are:

- Pre-school early education programmes
- Child care centres, including groups run by parents
- Home-based care.

12. The paper starts with an overview of recent trends in ECEC across OECD countries. This is followed by a synthesis of evidence on the impacts of ECEC on cognitive and social and emotional outcomes during the schooling years and later into adulthood. The paper considers variations across programmes and effect sizes over time. This section ends by identifying key gaps in the existing empirical evidence base.

13. Drawing on the previous work of the OECD, the OECD's ECEC Network and the European Commission, the third section outlines a conceptual framework that distils the relationship between child outcomes and the likely key determinants of these outcomes. The latter includes the level and type of ECEC, the child's home learning environment and the child's individual characteristics.

14. The paper then considers how international comparative data on early learning outcomes could add value, for policy makers and advisors, ECEC practitioners and parents. The next section presents a proposal to develop international comparative data on early learning outcomes, including the policy and research questions this a study would seek to answer, guiding principles for developing such data, and potential risks and mitigation strategies.

15. The next section canvases existing assessment measures of early learning, showing that such assessment is possible, even if some adaptation may be needed for an international comparative study.¹

16. The final sections of the paper identify:

- Links to other relevant OECD data work
- Indicative costs, and
- A proposed timeline and next steps.

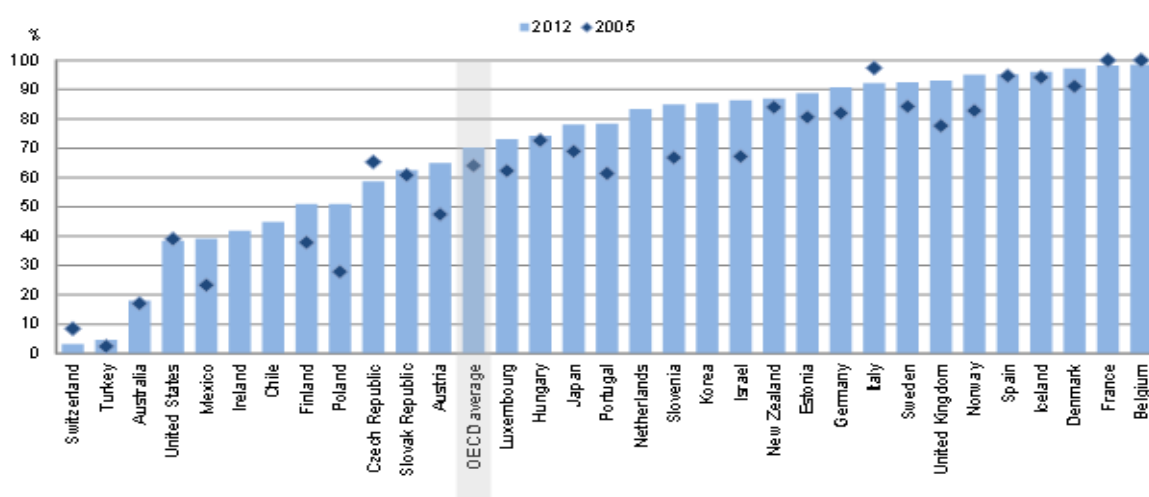
Trends in ECEC

17. Many countries have increased their efforts and progress on lifting children's participation in ECEC, including children from disadvantaged backgrounds. As participation rates have risen, increasing attention is being placed on the quality of participation, through measures to improve the quality of staffing, curriculum, and monitoring and data collection.

Participation rates and investment in ECEC are continuing to increase

18. Participation in ECEC has increased over the last decade in nearly all OECD countries, particularly amongst children aged three and younger. The following graph sets out relative enrolment rates for three year-olds.

Figure 1 Enrolment rates in education at age 3 (2005, 2012)



Source: Education at a Glance 2014, Chart C2.1.

19. These increases reflect increasing rates of women's labour market participation, as well as greater awareness of the benefits that high quality ECEC can provide to young children. However, despite this growth in participation across countries, demand still outstrips supply in some jurisdictions (OECD, 2014).

Improvements are being pursued in the quality of staffing

20. Many countries have taken action to improve the quality of ECEC, including:

- Raising staff qualification levels, often to align with those for primary school teachers
- Strengthening the content and standards of pre-primary initial teacher education programmes
- Improving the recruitment, retention and diversity of their ECEC workforces, through measures such as improved pay and other conditions.
- Continuously up-skilling the existing workforce.

21. Notwithstanding the above actions, a number of countries still report concerns regarding the quality of their ECEC workforces, and the overall status of this sector.

Curricula are also being redesigned

22. Almost all OECD countries have a curriculum or learning standards for ECEC provision from age three to compulsory schooling. In some cases, the curriculum expresses broad frameworks, which are then translated into greater detail at the regional, provincial or local level. Other jurisdictions have developed greater specificity on the learning outcomes children should gain from ECEC, i.e. what children should know and be able to do.

23. In recent years, curricula or learning standards have been embedded within a life-cycle or lifelong learning approach, and a growing number of countries and regions have started to frame continuous child development from early childhood through to older ages, including to the final year of schooling.

And monitoring and data collection is increasing

24. A number of countries have expanded the type of monitoring and data collection undertaken. As in the schooling sector, some of this expansion is related to a trend towards increasing devolution of responsibilities to regional, state and local levels (Morris, 2011). In addition, national ministries have increased the requirements on local government to actively manage ECEC provision and have provided a greater array of frameworks and support measures to assist in this.

25. Despite this greater focus on monitoring, few jurisdictions have confidence that their definitions of quality and monitoring regimes are resulting in on-going improvements in the quality and effectiveness of ECEC provision (OECD, forthcoming).

What type of ECEC provision can really make a difference?

26. An analysis of quantitative evidence on the relationship between ECEC and later outcomes enables some conclusions to be reached, but many questions remain unanswered.⁴ The evidence shows that:

- ECEC can have enduring positive effects, although effect sizes vary significantly
- These impacts are greatest for children from disadvantaged backgrounds, including for children with special learning needs in the early years of childhood
- The benefits of ECEC may be reducing over time, as participation rises
- Evidence on the factors that drive effectiveness in ECEC is weak.

ECEC can have enduring, positive effects

27. Most evaluations of ECEC interventions and programmes have found positive impacts on children's capabilities at entry to school, for both cognitive and social and emotional development. Measures of aggressive behaviour have been found to be lower amongst children that have attended ECEC and measures of sociability and concentration have also been found to be higher (Sylva et al. 2004).

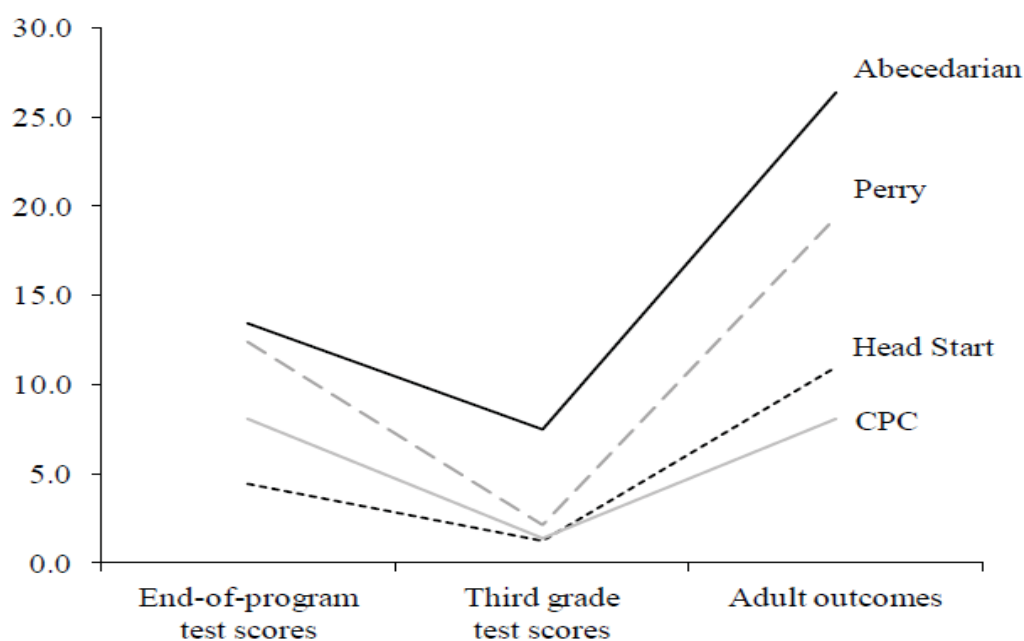
⁴ For a fuller description of the evidence on the impacts of ECEC see: Executive Office of the President of the United States. (2014). *The Economics of Early Childhood Investments*; European Commission. (2014). *Study on the effective use of early childhood education and care in preventing early school leaving*, Annex 1. Literature Review.

28. There is wide variability of effect sizes across and within programmes, including some studies which have found no positive effects of ECEC at all. (Driessen, 2004). In addition, a few studies have found negative effects in some settings, such as the impacts of extremely high levels of child minder care (Sylva et al. 2004) and for children in childcare at ages 0-3 years (Felfe and Lalive, 2011).

29. Most studies of the impact of ECEC on both social and emotional and cognitive abilities also confirm a fade-out effect in the early years of schooling. The most pronounced fade-out results arise in recent evaluations of Head Start⁵, which shows that by the end of the third grade Head Start children were academically indistinguishable from their peers who had not participated in the programme (Puma et al., 2012).

30. Positive effects of ECEC, however, tend to re-emerge in later years of schooling and endure into adulthood. An analysis of experimental, longitudinal studies in the United States on the effects of ECEC found positive benefits in late adolescence and into adulthood, as shown in the following diagram.

Figure 2 Predicted Percentage Effects on Adult Earnings of Early Childhood Programmes, Based on Test Scores versus Adult Outcomes (Bartik, 2014)



31. The reasons for the re-emergence of positive effects are not well understood. Bartik notes that this fading and re-emergence of effects could be due to non-cognitive skills, which are important to adult earnings and are not always measured using standardised tests. Social skills and character skills, he suggests, are at least as important as cognitive skills in making a worker more employable and more productive (Bartik, 2014).

32. From this analysis, Barkik concluded that the initial test score effects are better predictions of adult earnings effects than the faded test score.

⁵ Head Start began in the United States in 1965 and provides ECEC to 3- and 4-year olds from low income households.

The impacts are greatest for disadvantaged children

33. The relative impacts of ECEC on disadvantaged children are positive. Studies have found that children from low income households benefit more in terms of learning gains from ECEC than children from higher income households (Duncan and Soujourner, 2012). However, these relative gains do not fully ameliorate the effects of disadvantage (Sylva et al., 2004).

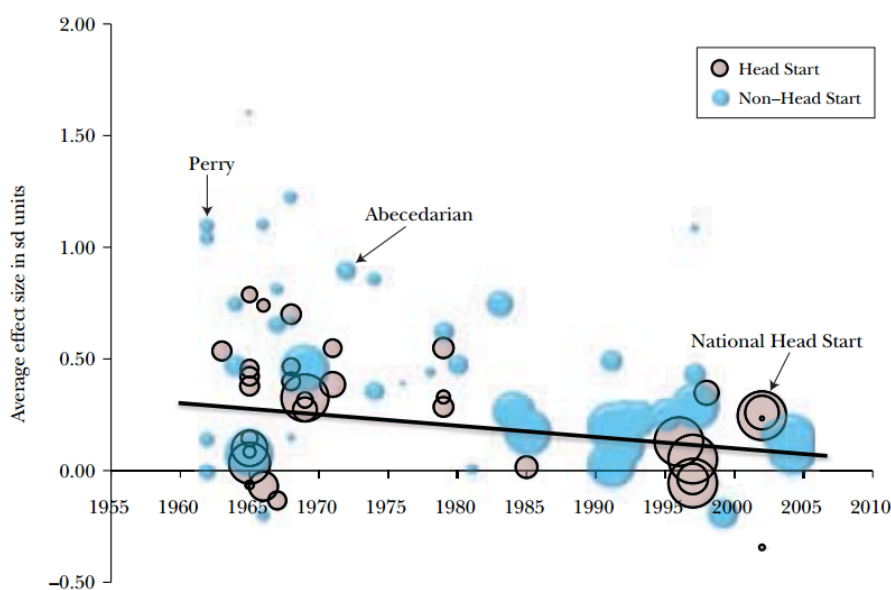
34. ECEC has also been found to be of particular benefit to children at risk of developing learning difficulties. For example, in the EPPE⁶ study in the United Kingdom, one in three children faced risks at the start of pre-school, and this reduced to one in five by the time these children started school (Sylva et al., 2004).

But the benefits of ECEC may be reducing over time

35. A meta-analysis of 84 ECEC programmes found that the magnitude of measured effects of these programmes had declined over time, as illustrated in the graph below (Duncan and Magnuson, 2013).

Figure 3

Average Impact of Early Child Care Programs at End of Treatment
(standard deviation units)



(Duncan and Magnuson, 2013)

⁶ The Effective Provision of Pre-school Education Project (EPPE) started in the United Kingdom in 1996 and has followed the impact of ECEC on children until the age of 14.

36. This decline may reflect a combination of factors including:

- Early programmes such as the High/Scope Perry Preschool Project⁷ and the North Carolina Abecedarian Project⁸ were designed and implemented with small cohorts of children, enabling the quality of provision to be closely controlled and monitored (Baker, 2011);
- Children in control groups are now more likely to attend some form of ECEC than the control groups in the 1960s and 1970s, when the Perry and Abecedarian Projects were run respectively⁹;
- Children's home learning environments may be improving, as parents' education levels increase (Duncan and Magnuson, 2013) and as parents spend more time with their children (Ramey and Ramey, 2010);
- The rapid expansion of ECEC provision and the consequent need to recruit a larger pool of ECEC staff, which may have resulted in lower quality teaching in some settings.

37. While the factors or combination of factors pertinent to each programme will undoubtedly vary, the evidence suggests that further lifting ECEC participation rates will not necessarily lead to the gains that early programme evaluations suggested were possible.

Evidence on the factors that support good learning outcomes is weak

38. As in the schooling sector, the relationship between structural aspects of ECEC such as class size and children's development has been found to be weak (Mashburn et al. 2008). Even factors such as teacher qualifications and management and leadership do not always show a clear relationship with children's development (Bradshaw et al., 2014; Gialamas et al., 2014). Similarly, programmes with the same structural design do not produce the same results. In Head Start, for example, varying results are found in centres that use the same curriculum, the same proportion of educated and certificated staff, and the same class size (Walters, 2014).

39. The above findings may relate to the quality of the proxy being used, rather than conclusively demonstrating that such factors do not play a part in determining quality and effectiveness. This evidence does, however, suggest that the drivers of quality in ECEC are likely to be complex rather than simplistic.

Are there significant data gaps?

40. The significance of existing data gaps depends on the policy and practice questions that remain unanswered. The key questions on the benefits that may be gained from ECEC include:

- Does ECEC make a difference?

⁷ The High/Scope Perry Pre-school Project was run in Michigan from 1962 to 1967 for low income African American 3 and 4-year-olds.

⁸ The North Carolina Abecedarian Project operated initially from 1972 to 1977, for low income children from 6 weeks after birth to 5 years of age.

⁹ A recent analysis of a Head Start control group found that 60% of these children had attended ECEC (Puma et al., 2012).

Terms of Reference

- For whom?
- In what domains?
- By how much?
- For how long, and
- Under what circumstances?

41. The empirical evidence answers some of these questions, at least partially, but not all. Current gaps in the empirical data include:

- Recency of findings. ECEC sectors and families are dynamic and findings from five to ten years ago may not be valid today. Without recent data, it is difficult to confidently reach conclusions and recommendations on how to improve current policies and practices
- Applicability in different systems and cultural settings. Most empirical data is American-based. While some empirical data on the impacts of ECEC exists in other OECD countries, this is not the case across all member countries
- Outcomes across different groups of children. While the key American studies focused on low income, mostly African-American children, there are few studies that cover all socio-economic groups or key groups of children, such as those from diverse ethnic and migrant backgrounds and those with special needs
- Parents and families are a critical ingredient in child development, but little information exists on what helps and what hinders families in providing a rich learning environment for their child. As the EPPE study revealed, what parents do with their child in the early years is much more important than who the parents are (Sylva et al., 2004)
- Specificity on the factors within ECEC settings that drive quality. There is more information on indicators of structural quality, such as class size, than on process quality, i.e. the quality of the interactions between ECEC staff and individual children. As noted earlier, structural indicators of quality do not appear to have high predictive validity for children's learning and development
- System-wide performance information. Most evaluations are of particular programmes rather than benchmarking the performance of ECEC systems as a whole, including the cost of those programmes to parents and taxpayers
- Cross-system comparisons. The lack of a comparative measure of early learning on outcomes means countries are not able to see what is possible to achieve in this sector. Nor can they learn from each other on how to improve the effectiveness, equity and efficiency of their ECEC systems.

A conceptual framework on early learning outcomes

42. The conceptual model below describes the relationships between particular outcomes and the key determinants of those outcomes. It attempts to identify the particular learning outcomes of interest, the key elements of ECEC that influence those outcomes, as well as other relevant determinants of children's learning outcomes. The latter includes, for example, the child's home environment as well as the child's individual characteristics. This conceptual model – or framework – then provides a structure for developing the key policy questions that need to be addressed and for assessing the significance of any gaps in the existing data and wider evidence base.

43. This section sets out how child outcomes were identified as a priority for the OECD's ECEC work programme, foundational work already completed for this work stream and then a draft conceptual framework on early learning outcomes, for discussion and feedback.

Child outcomes were included in the OECD's programme of work and budget for 2015/16

44. A data mapping exercise in 2012 raised a gap in relation to international data on child outcomes and child development¹⁰. A data development strategy was then formulated in conjunction with the OECD's ECEC Network. The strategy was based on the policy questions identified by countries as most relevant to them in the area of ECEC. Several countries raised interest in child outcomes data and in being able to show 'return on investment' for expenditure on ECEC.

45. The data development strategy¹¹ set out a timeline to identify available instruments for data collection in 2014, explore the development of new instruments in 2015 and pilot data collection in 2016. Consequently, the Programme of Work and Budget 2015/16 includes, amongst other deliverables, the development of a conceptual framework on child outcomes in the early years ([EDU/EDPC\(2014\)24](#)).

46. As a foundation for the OECD's work on child outcomes, the ECEC Network produced a paper on "Common Understandings", to provide a common language about young children's learning and development¹². The paper highlights the:

- Importance of early childhood in terms of later outcomes
- Critical importance of placing the child at the centre, i.e. a holistic view of the individual child and his/her needs, strengths, interests, language and culture
- Diversity of children's cultures, backgrounds and needs
- Critical role of parents
- Positive impact high quality ECEC programmes and services can have for children, which should be designed to integrate care, socialisation and learning.

¹⁰ Revised project proposal of new policy output on early learning and development, available on: [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=EDU/EDPC/ECEC\(2012\)3/REV1/ANN2&docLanguage=En](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=EDU/EDPC/ECEC(2012)3/REV1/ANN2&docLanguage=En)

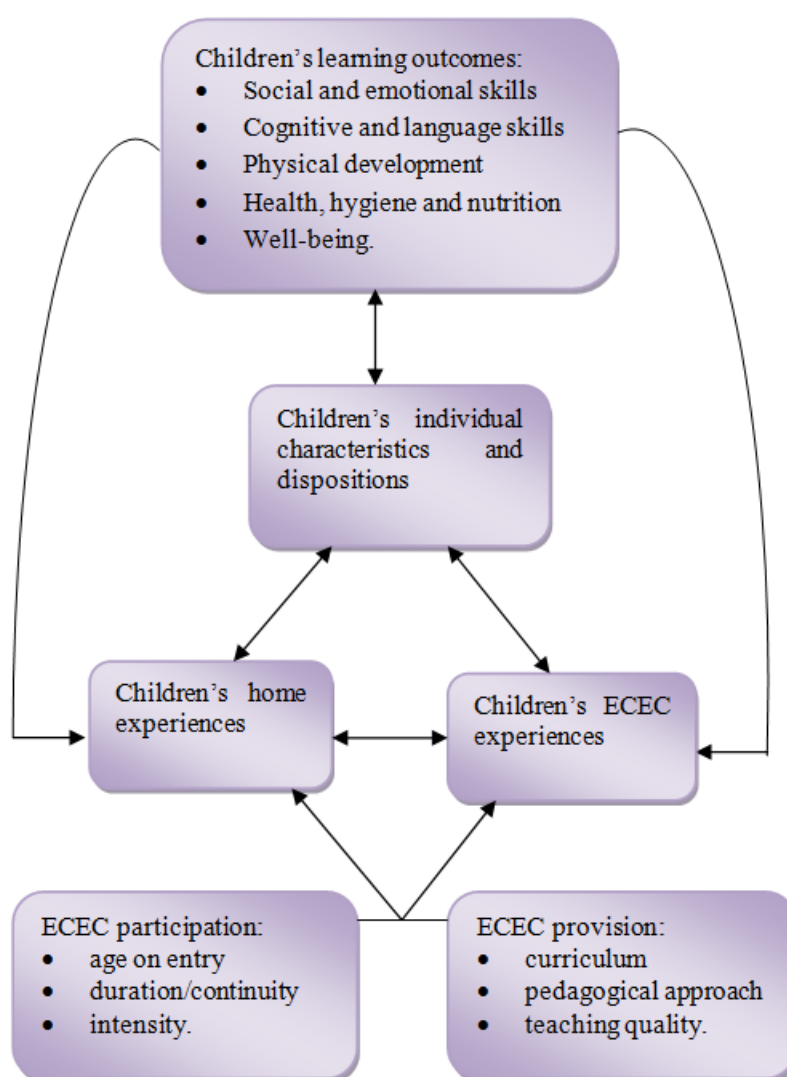
¹¹ The Draft Proposal on the Data Development Strategy of the OECD Network on Early Childhood Education and Care, available on: [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=EDU/EDPC/ECEC\(2013\)1&docLanguage=En](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=EDU/EDPC/ECEC(2013)1&docLanguage=En)

¹² See 2

A conceptual model of the relationship between ECEC and child outcomes

47. In developing a conceptual model of early learning outcomes, the following framework draws on the work of the ECEC Network, the European Commission's Thematic Working Group on ECEC¹³ and the conceptual model developed by Vandell and Wolfe on the relationships impacting early childhood development outcomes (Vandell and Wolfe, 2000). The model illustrated below demonstrates the relationship between child outcomes, the child's own characteristics, the child's home environment and the child's ECEC experiences.

Figure 4



¹³ The Thematic Working Group reported in October 2014 with a "Proposal for key principles of a Quality Framework for Early Childhood Education and Care". (European Commission, 2014).

Terms of Reference

48. Social ecological models are also relevant in the context of an international assessment of early learning. For example, Bronfenbrenner's ecological systems theory includes the effects of nationwide cultural and organisational aspects on the human's development (Bronfenbrenner, 1979). His framework states that four ecological systems can affect a child's growth:

- **Microsystem:** The microsystem is the closest environment to an individual. It comprises relationships and interactions with immediate surroundings, such as parents or a child's teacher.
- **Mesosystem:** The mesosystem embraces interactions among various stakeholders of the microsystem, e.g. the relationship between a child's parent and the child's school.
- **Exosystem:** The exosystem does not involve a child directly, but covers aspects of structures within the microsystem, such as familial financial difficulties.
- **Macrosystem:** The macrosystem is the furthest environment to an individual, and it embraces the overall patterns of ideology of a given society, e.g. laws, life-styles, customs, beliefs... This framework therefore conceptualizes how children's development can vary across countries due to differences in cultures and systems.

49. The sections below expand on Figure 4's key determinants:

- Children's learning outcomes
- Children's individual characteristics and dispositions
- Children's home experiences
- ECEC participation, and
- ECEC provision.

50. For each area, the paper provides a brief synopsis on the existing evidence, sets out likely variables of relevance and probable key policy and research questions.

Children's learning outcomes

51. As shown by the research evidence earlier in this paper, children's early learning development has significant impacts on their later educational success, as well as other outcomes. Children who do not make sufficient progress in the first few years of their childhood face huge barriers in catching up sufficiently to achieve success at school (Naudeau et al, 2011). Thus, it is critical that children's development is monitored both at an individual level and at a system level. And although each child is unique, the basic patterns, or principles, of growth and development are universal, predictable, and orderly (European Commission, 2012).

52. While children's development is highly inter-related across domains and skills, there are a range of early learning outcomes that can be assessed. These include:

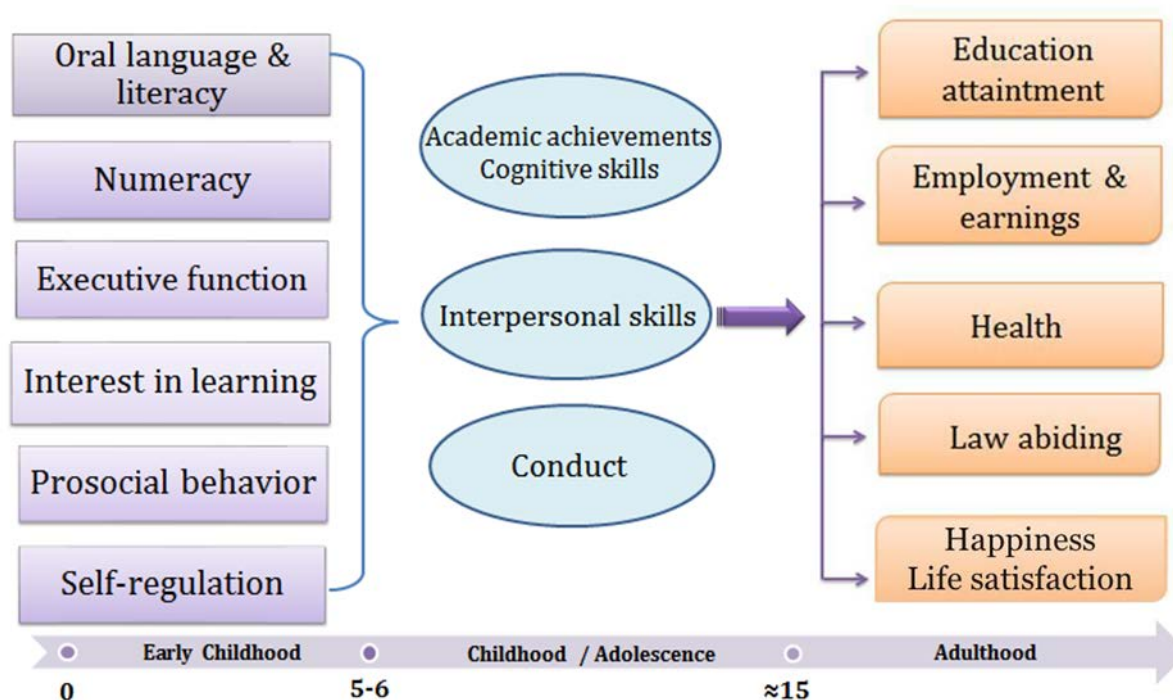
- Social and emotional skills
- Communication and language skills
- Other cognitive skills
- Physical development
- Health, hygiene and nutrition
- Well-being.

53. The domains used in any assessment should reflect the purposes for that assessment and what the information will be used for. The context in which the assessment is taking place is also important. For example, if almost all children across jurisdictions are healthy and are regularly monitored by health professionals, assessing children's development in this domain is unlikely to add much value.

54. Evidence from longitudinal studies that have investigated the relationship between children's early learning and later outcomes in adulthood suggests that the key predictive domains are likely to include:

- Oral language
- Literacy
- Numeracy
- Executive function¹⁴
- Interest in learning
- Pro-social behaviour
- Self-regulation.

55. These are illustrated in the following diagram.



56. However, this list is not exhaustive and other early domains might be of interest, such as self-esteem. Moreover, other lifetime outcomes could be included in the study, such as social and democratic participation.

¹⁴ Executive function refers to planning, focusing attention, remembering instructions and juggling multiple tasks successfully. Executive function relies on the use of working memory, mental flexibility and self-control.

Terms of Reference

57. The key policy and research questions in terms of the outcomes or domains to be assessed are:

- What is most important to know in terms of how well are children developing in the early years?
- What domains are of greatest interest?
- How areas of relative strength and areas of concern will be identified?

Children's individual characteristics and dispositions

58. Every child is unique and has his or her own personality and particular disposition, likes and dislikes, abilities and challenges. Thus, children start with endowments of cognitive potential and temperament, which are determined by genetics, although these can and do change over time

59. Children's other characteristics also affect their early learning experiences, such as their socio-economic status, ethnicity, migration status, region or locality, and gender.

60. Learning disabilities and other special needs are also crucial to understand. For some of these needs, ECEC can be highly effective in early identification and addressing particular learning challenges.

61. Key policy and research questions are likely to include:

- How well does the ECEC system deliver for different groups of children?
- How much variation is there in these results?
- What combination of variables correlates with better outcomes for different groups of children?

Children's home experiences

62. Children's families provide their first learning environment. The quality of the child's home learning environment plays an important role in shaping children's cognitive and social and emotional development. Thus, the impact of ECEC provision must be considered in the context of the nature and quality of the child's home learning environment.

63. The EPPE findings indicate that the quality of the home learning environment is more important for children's early development than parental occupation, education and income. As the EPPE study concluded, what parents do is more important than who they are. In addition, the EPPE project found that children whose parents engaged regularly in home learning activities were less likely to be at risk for special education needs (Sylva et al., 2004). Children's families determine what type of ECEC the child participates in, as well as the age of starting ECEC and whether participation is continuous.

64. The relationship between a child's family and ECEC providers is also important. ECEC programmes that foster the participation of parents and involve them in pedagogical decision-making and in their children's learning can enhance children's early development (Lazzari & Vandenbroeck, 2012).

65. Key policy and research questions are:

- What knowledge do families have on how to support their child's development?
- What helps or hinders parents in supporting their child's development?
- What are the relative strengths of children's home learning environments?
- To what extent do ECEC services enhance the impact of the home learning environment?

ECEC participation

Terms of Reference

66. The EPPE project found that both an earlier age of entry and longer duration of ECEC participation correlated with better gains for children, especially in developing cognitive skills. The study concluded that children who start pre-school from the age of three have a significant advantage over those children who attend only one year of pre-school before entry into primary school (Sylva et al., 2004).

67. Findings are mixed on whether full-time or part-time provision is most beneficial. For example, the EPPE study found that full time attendance led to no better gains than part-time, whereas full-day Head Start programmes have been found to boost cognitive skills (Walters, 2014).

68. The EPPE project also found that disadvantaged children benefited more when they were in a setting that included a mix of children from different backgrounds. This finding has not been tested in the other studies referred to in this paper, as most of these programmes targeted disadvantaged children.

69. Policy and research questions would focus on:

- What participation starting age, duration and intensity are most beneficial to children?
- What works best for particular groups of children?

ECEC provision

70. The quality of the ECEC services provided and the approach taken can significantly impact on the experiences of children and the benefits they receive. As noted earlier, structural aspects of ECEC provision are not reliable predictors of effectiveness.

71. The EPPE project found a set of characteristics that were positively associated with effectiveness:

- The quality of adult-child verbal interactions
- A mix of adult and child initiated activities
- Qualified staff, with knowledge of the curriculum and child development
- High parental engagement, and
- Support for children's behaviour management.

72. There is great diversity in the approaches taken in ECEC delivery, and across indicators of both structural and process quality. This includes whether the provision is based at a centre, a school or home-based; whether it is delivered by qualified teaching professionals, untrained staff or parents; class and group sizes; the physical space available to children; and the monitoring regimes in place. Pedagogical approaches also differ, from the child-centred model of Reggio Emilia to more instructional delivery. There are also curriculum differences across countries and differences within a country in how the curriculum is delivered.

73. Relevant policy and research questions may include:

- What types of provision support different groups of children best?
- What structural indicators of quality affect child outcomes, e.g. staff qualifications, teacher salaries, staff/child ratios, group size, group mix, physical space?
- To what extent does the curriculum affect learning?
- What pedagogical approaches best support learning for different groups of children, and in what domains?

How would international comparative data add value?

74. The purpose of any international comparative study should be to help countries improve the performance of their systems, to provide better outcomes for citizens and better value for money. Comparative data can show which systems are performing best, in what domains and for which groups of students. It would also provide insights on how such performance has been achieved. Thus, internationally comparable data would enable countries to compare the relative strengths and areas for development in their own ECEC systems with those in other jurisdictions.

75. The measurement issues involved in an assessment of early learning outcomes are necessarily complex. By pooling expertise, countries will get a more robust set of measures than they would likely achieve by working individually. The overall value of this data will be the extent to which it is able to help countries to improve the experiences and outcomes of future cohorts of children.

76. The following section outlines the likely benefits international comparative data would provide for policy makers and advisors, for ECEC practitioners, for parents and for other stakeholders.

For policy makers and policy advisors

77. International comparative data would show policy makers and policy advisors what is possible to achieve from ECEC, in terms of effectiveness, equity and efficiency, and would enable them to position their own jurisdiction against best practice. Within their own system, the data would provide insights on:

- How well young children are positioned to succeed, at an early point in their learning pathways
- How much variation there is within systems and across different types of provision
- How well disadvantaged groups of children are faring, and
- The individual, institutional and systemic factors that relate to learning outcomes.

78. Other areas of likely interest that also may be possible, even if at a later stage in the study, are:

- The participation experiences that support children most, such as age of entry to ECEC, and the intensity, duration and continuity of participation
- The types of ECEC provision that are most helpful to different groups of children and in different domains
- The types of engagement with and support to parents that are most effective in advancing children's learning
- The relative impacts of structural factors, such as staff/child ratios and staff qualifications
- The influence of pedagogical approaches and indicators of process quality.

79. The information from this study will also assist decision-makers to better understand the further contribution that their ECEC systems can make for improving children's learning, in relation to the possible further contribution that early primary schooling can make.

80. The study will enable information on early outcomes to be available in a timely manner, so that changes in policies, such as in funding and regulations, may be made for the benefit of the next cohort of early learners.

81. In time, the information can also provide information on the trajectory between early learning outcomes and those at age 15, as measured by PISA. In this way, countries can have an earlier and more specific indication of how to lift the skills and other capabilities of its young people.

Terms of Reference

82. Data gathering on an international basis also provides benefits in terms of economies of scale. The development of instruments to match the priorities and needs of participating countries means that countries obtain the measures most relevant to them and that the development costs of these instruments are shared.

For ECEC practitioners

83. ECEC practitioners are commonly conducting assessments of children's learning and development as an integral part of their teaching, but generally not as part of standardized tests (Barnett, forthcoming). This assessment information is largely not collated and thus cannot be used to inform policy or system wide practice decisions.

84. Having access to national and international comparative data on early learning outcomes will provide a valuable resource for ECEC staff. This information can support ECEC practitioners to reflect on their practice, and see both the strengths of their current practices and possible areas of improvement. It will also provide a body of evidence on what children are capable of achieving.

85. While the opportunities for ECEC practitioners to engage in professional learning will vary across jurisdictions and settings, using this data to reflect on the learning and development of the children the teacher is working with and the teaching and learning strategies being used is likely to have a positive impact on children's learning outcomes (Timperley, 2008).

For Parents

86. Parents are a critical determinant of children's early learning, both in terms of the quality of the home learning environment and on whether the child participates in ECEC and the nature of this participation. Parents also hold critical information on the individual dispositions of their child and on the effects of ECEC on their child.

87. For many parents, their choices on their child's ECEC participation are constrained by lack of availability, cost, access difficulties such as transport, and concerns about quality, including the likely responsiveness of an ECEC provider to them and to their child. Information on early learning outcomes, however, could provide them with reliable information on:

- What practical activities they can undertake with their child that will make a significant difference to their child's learning and make the most from their ECEC and schooling experiences
- What age it would be beneficial to enrol their child in an ECEC setting and what is likely to be best in terms of intensity, duration and continuity
- What kinds of capabilities their child should be building, in both social and emotional and cognitive domains.

For other stakeholders

88. There are a number of other stakeholders who would also benefit from such comparative data. These include:

- Leaders and practitioners in early schooling
- Owners of ECEC centres
- Funders of ECEC provision

Terms of Reference

- Researchers and advocates in the fields of early education and child and family policies
- Media and other public commentators.

89. For each, the data would provide greater insight into the strengths and weaknesses of their country's ECEC system, the expectations they could have of their own systems, and what their country may be able to learn from other systems.

Developing a proposal

90. The purpose of this paper is to assess what contribution international comparative data on early learning outcomes would add. The paper is intended to assist countries to decide whether they wish to be involved in scoping such an international assessment and, at a later point, potentially participate in such an assessment.

Decision to enter a scoping phase

91. Questions countries may use to reach a decision on whether to enter a scoping phase in collaboration with other participating countries and the OECD are:

- Would information on the overall quality and impact of your ECEC system be helpful in improving ECEC policies and practices, as well as early schooling?
- Would benchmark information be helpful to monitor ECEC system performance over time?
- Would information on the performance of other countries' ECEC systems be useful to see the relative strengths across different systems and approaches that are working well?
- Are there particular groups of children or particular domains that should be highlighted?
- Would information from parents provide valuable insights into children's home learning environments and on how well ECEC practitioners work with parents?
- Is input from teaching practitioners also critical?

92. Two webinars are planned for April and June 2015 (see later section on proposed timeline and next steps). The webinars will enable countries to find out more about the study, ask questions and to further consider the value this project could add to their efforts to improve children's early learning.

Guiding principles

93. In developing the work further there are a number of guiding principles that should apply, such as:

- the work should be developed and judged in terms of how well it adheres to and responds to the policy questions participating countries have agreed upon, i.e. policy relevance
- the information produced will enable countries to better understand their ECEC systems and how to improve them, i.e. practicability
- the information must be reliable, valid and comparable

Terms of Reference

- the well-being of all children in the study must be of utmost priority
- the impacts on ECEC staff and parents must not be onerous and distract from their roles in caring for children.

Scoping the study

94. Countries that participate in the scoping phase will be able to determine the overall intent, objectives and shape of the study. Outputs from the scoping phase will include:

- A clear statement of agreed policy objectives. These will set out what countries are trying to achieve from the study, such as improvements to ECEC policies and practices and the overall performance of their ECEC systems
- Agreement on the key policy questions that will be covered by the study. This will include the domains to be covered, including the relative weightings of social and emotional and cognitive domains; the extent to which home learning environments and the perspectives of parents will be included; and differentiation across types of provision and groups of children
- A request for proposals, which incorporates the policy objectives and questions, but also includes other requirements, such as those relating to quality assurance and timeframes
- Selection of a preferred contractor to develop an assessment framework and instrument design, and to implement a pilot.

95. In clarifying the policy objectives, a key decision will be whether countries main interest is in:

- How well all children are positioned, including those with no ECEC history, or
- How well children who have experienced ECEC are positioned.

96. Focusing on all children would provide countries with information on children with a range of ECEC experiences, and point to what might be needed to support different groups of children, including those who have not participated in ECEC. It would also provide a comparison between those children who have participated in ECEC and those who have not.

97. Focusing on children who have participated in ECEC will provide countries with greater information on the differential impact of different ECEC services and approaches. This would help to identify the relative strengths and areas for improvement within ECEC systems.

Risk management

98. There are a number of risks inherent in developing any new assessment measure, which will need to be managed carefully. The OECD has significant experience in successfully developing and implementing international assessments across diverse countries. While an assessment of early learning outcomes would present unique issues, the OECD's experiences position it well to undertake such an assessment and produce high value-add results for countries.

99. Risks that the project will manage are:

- Impacts on children. In considering any potential instrument and assessment process, any potential negative impacts on children would exclude the use of any measure. As noted in the

Terms of Reference

section below, there are a number of measures already in place that measure children's early learning development without negative impacts on these children

- Rigour, validity, reliability and comparability. The OECD will use its in-house expertise and external experts to ensure only valid and reliable instruments and assessment processes are used. As noted above, the organisation has significant experience in undertaking assessments across languages and in different cultural settings
- Burden on practitioners and/or parents. Countries are best placed to assess the level of time that would be acceptable for practitioners and for parents. Some countries may wish to engage relevant stakeholder groups to gain feedback on this
- Cost. The potential costs for this type of assessment are wide-ranging. Actual costs will be determined by countries' decisions on the items assessed and the assessment processes used.

Approaches already in place

100. A number of measures of early learning outcomes already exist, which may enable less costly and speedier development and implementation of an international assessment measure than would otherwise be the case.

101. The ECEC Network recommended a stocktake be undertaken of existing instruments that countries use or have used to measure child outcomes, and how the results from such assessments are used. In response to this request, the OECD commissioned Dr Steven Barnett from Rutgers University to complete a report on these issues.

102. Barnett concludes that there are a wide range of existing measures available, which may be able to be used as the basis for an international measure of early learning and development. Barnett investigated the domains covered by 20 separate measures and found that nearly all included measures of social and emotional skills, communication and language, other cognitive skills, and physical development, as illustrated in the following table.

<i>Domain</i>	<i>Proportion of instruments assessed that include the domain (%)</i>
Social/emotional skills	100
Communication and language	100
Other cognitive skills	90
Physical development	100

103. Thus, the development work involved in shaping international comparative measures of early learning could be streamlined if existing measures can be used, even if modification is required.

104. The report on existing measures of early learning outcomes notes differences between measures in terms of:

- The purposes they were developed and have been used, i.e. formative, summative or both

Terms of Reference

- The methodologies employed
- Their comprehensiveness across the domains covered
- Applicability across countries, languages and cultures
- Relative technical robustness, and
- The costs in applying the instrument, such as training and moderation, as well as the time required of teachers, parents and children.

105. The report also notes that in selecting or developing any measure of early learning outcomes, it will be critical to:

- Measure what matters, i.e. those aspects that are important and of concern to policy makers and the public
- Measure well, i.e. valid, reliable, fair and age and developmentally appropriate
- Ensure assessments are practical and affordable, and
- Obtain results that are comparable within and across countries and over time.

106. The report concludes that if any existing measure is selected as a basis for an international comparative assessment, it will likely need some adjustments to enable it to work for the specific purposes required and to be appropriate across diverse country settings.

Early learning assessment for developing countries

107. A recent initiative to assist developing countries to improve early learning outcomes has been developed by UNESCO, UNICEF, the World Bank and the World Health Organisation. The Measuring Early Learning Quality and Outcomes Project provides open-source, freely-accessible measures of child development and learning and the quality of early learning environments. This resource is intended to help developing countries to address poor learning outcomes, through enabling the monitoring of progress towards national and global goals by identifying children's competencies and areas of need.

108. While developing countries have different needs from OECD countries in terms of their assessment measures, standards and processes, it would be useful in developing a new tool for assessing early learning outcomes if there was a means for developing countries to progress to an OECD country measure, in the same way PISA for development is intending to achieve.

Links to other OECD assessments and surveys

109. There are several relevant OECD comparative assessment measures and surveys that are relevant to an assessment of early learning outcomes:

- The ECEC staff survey
- The Education for 2030 project
- The Education and Social Progress longitudinal project
- PISA.

110. Each of these is discussed below.

The ECEC staff survey

111. A number of countries have indicated their interest in the ECEC staff survey and are currently contributing towards shaping the design of the survey. The survey will provide in-depth information on ECEC staff's experiences, views, beliefs and pedagogical practices. Thus, it will provide a rich and comparative information base on the context and capabilities within ECEC services.

112. Some countries have indicated an interest in participating in both the early learning outcomes assessment and the staff survey. Participating in both will provide greater insights on both the performance of a country's ECEC system and the context for these results. It will be possible for countries to participate in both projects and the current draft timelines have both the early learning assessment and the ECEC staff survey being implemented at the same time. Depending on the final design of each, it may be possible to achieve some savings for countries who participate in both, by eliminating duplication in the information collected.

113. At the same time, the outcomes project and the ECEC staff survey can be run entirely separately. Thus, countries are able to opt into one, and not the other.

Education and Social Progress

114. The Education and Social Progress Project (ESP) is a longitudinal study of skills development in cities. It will measure the social and emotional skills of children from Grade 1 (approximately age 6) until early adulthood. Under the current timeline, the first assessment at Grade 1 is scheduled to occur in 2020.

115. The project has investigated the emotional and social competencies that support later success across a number of domains, and is currently moving into a feasibility phase. The project provides a significant body of work to inform the development of the social and emotional outcomes that could be assessed through the early learning outcomes project. Clearly, there is an overlap between the two assessments and some degree of alignment should be achieved to enhance the insights for countries that participate in both.

Education for 2030

116. The Education for 2030 project aims to support countries to place a future-focused, innovative lens on the kinds of skills, competencies and values needed to respond to a dynamic, changing world in 2030, at global, national and regional levels. The project is intended to support countries to think and re-think new models of curriculum, pedagogy, teacher education and assessment tools.

117. While still at an early stage of development, 2030 will provide insights on the types of education policies and practices that can help all students to be successful in 2030 and beyond. Also, the project aims to analyse change processes with a bottom-up dimension, inclusive of students, teachers and industry.

PISA

118. There are two related points of alignment with PISA. The first is the development of global competencies for inclusion in PISA 2018. The framework development for the global competencies is underway, and was discussed at the PISA Governing Board meeting in March 2015. Field trials are intended to take place in 2017, for inclusion in the main PISA assessment in 2018. This work potentially

links to both the 2030 and ESP projects, and will also inform the domains selected for early learning assessments.

119. The second relationship with PISA is to enable countries to link early learning outcomes to the capabilities of the same cohort of students at age 15. This will give countries greater information on what must be done within the schooling years, to give students a better chance of developing the skills they need and achieving success. If the first early learning outcomes assessment runs in 2017, depending on the age or stage selected for assessment, some of this cohort of students will likely undertake PISA in 2027.

Indicative costs

120. The costs for this type of assessment can vary widely, because of decisions on:

- The breadth of domains assessed
 - Whether an existing instrument is used or modified, or an entirely new set of measures are developed
 - The extent to which independent observation is used versus tools that teachers can use
 - The training requirements for local assessors, both independent observers and teachers
- Whether input from parents is through questionnaires, interviews or a mix.

121. Until the project is scoped, it is not possible to estimate the likely costs for countries of participating. However, costs can be estimated for working with countries to finalise the objectives and scope of the study, selecting and engaging a contractor, managing the development or adjustment of assessment measures and processes, and overseeing a pilot. These would be the costs incurred by the OECD in 2015/16, exclusive of the fees and other costs of a contractor and the national costs of running a pilot.

122. To manage the uncertainties about the overall costs of this study, the OECD is asking countries to only make a financial contribution to the scoping phase at this stage. This contribution and involvement will not commit countries to further funding contributions or participating in the survey.

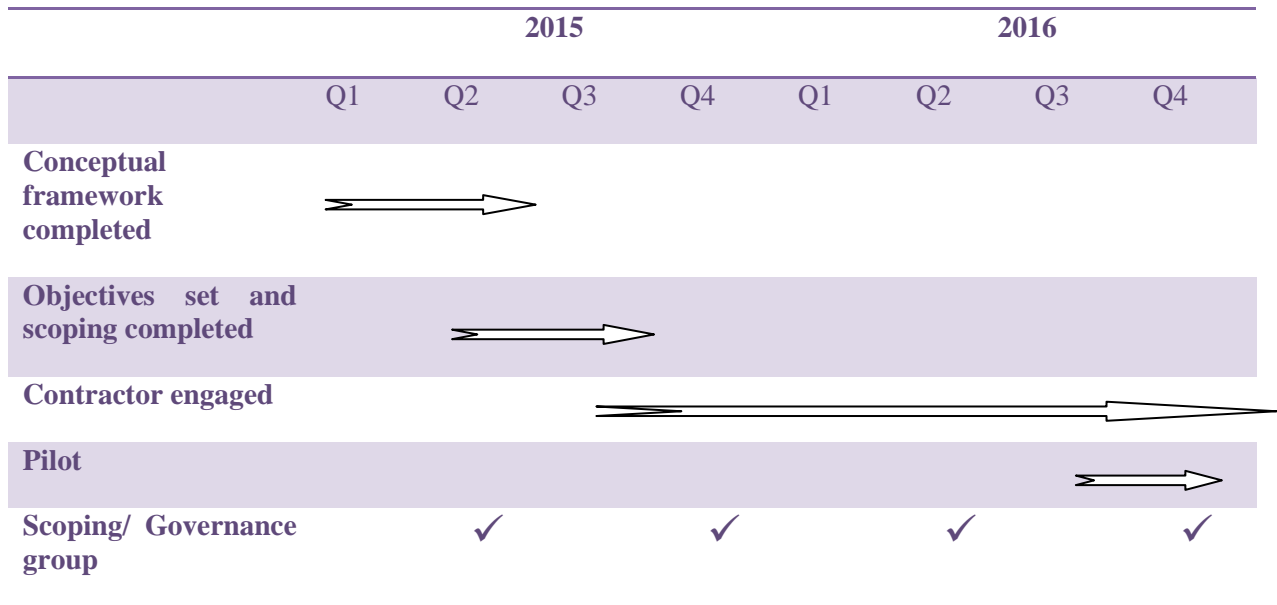
123. Countries will be asked to make a financial commitment, prior to participating in the scoping exercise, and these contributions will be required from July 2015.

124. At the point that accurate information is available on the likely costs of developing the assessment instrument, piloting the instrument and implementing the main study, this information will be made available to countries to enable countries to assess their involvement in any further stages. Note that the three main stages of the project are:

- Phase One: Scoping and selecting a preferred international contractor
- Phase Two: Developing the assessment instrument and overseeing a pilot
- Phase Three: Implementing the main study.

Proposed timeline and next steps

125. A proposed timeline to complete a pilot is as follows:



126. The next step in this project is the first meeting of interested countries on 8-9 July 2015, Paris. Agenda items will include objectives and draft scoping for the project. The latter will cover the key design features of the study, including the target group of children to be assessed and likely domains.

127. The European Commission has expressed its interest in collaborating with the OECD on this project. While participating countries will still determine the direction and shape of the study, the Commission's expertise on ECEC issues will be of great benefit as the project develops.

REFERENCES

- Adamson, P. (2008). The Child Care Transition, *UNICEF Report Card* (8)
- Barnett, S., S. Ayers and J. Fancis (forthcoming), Comprehensive Measures of Child Outcomes in Early Years, Report prepared for the 16th Meeting of the OECD Network on Early Childhood Education and Care, 18-19 November 2014, Berlin, Germany, OECD, Paris.
- Baker, M. (2011) ‘Innis Lecture: Universal early childhood interventions: What is the evidence base?’, *Canadian Journal of Economics*, 44.
- Bernal, R, and Keane, M.P. (2011). Child care choices and children’s cognitive achievement: The case of single mothers, *Journal of Labor Economics*, 29(3).
- Bartik, Timothy J. (2014). *From Preschool to Prosperity: The Economic Payoff to Early Childhood Education*
- Bradshaw, P., Lewis, G. and Hughes, T. (2014). Growing Up in Scotland: *Characteristics of pre-school provision and their association with child outcomes*, Scottish Government.
- Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press.
- Bull, R., Espy, K.A. and Wiebe, S.A. (2008). Short-term memory, working memory, and executive functioning in pre-schoolers: Longitudinal predictors of mathematical achievement at age 7 years, *Developmental neuropsychology*, 33(3).
- Caneiro, P., Crawford, C., and Goodman, A. (2007). The impact of early cognitive and non-cognitive skills on later outcomes.
- Clark, C.A., Pritchard, V.E., and Woodward, L.J. (2010). Preschool executive functioning abilities predict early mathematics achievement, *Developmental psychology*, 46(5).
- Cunha, F. and Heckman, J. J. (2010). Investing in Our Young People, *NBER Working Paper* (16201)
- Del Boca, D., Pasqua, S., Suardi, S. (2013). *Childcare, family characteristics and child outcomes: An analysis of Italian data*
- Driessen, G. (2004). A Large Scale Longitudinal Study of the Utilization and Effects of Early Childhood Education and Care in The Netherlands. *Early Child Development and Care*, 174(7-8)
- Duncan, G.J., Dowsett, C.J., Claessens, Al, Magnuson, K., Huston, A.C., Klebanov, P., and Japel, C. (2007). School readiness and later achievement, *Developmental psychology*, 43(6)
- Duncan, Greg J. and Katherine Magnuson. (2013). “Investing in Preschool Programs.” *Journal of Economic Perspectives*, 27(2).

Duncan, Greg J. and Murnane, Richard. (2011). "Introduction: The American Dream, Then and Now." In *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*. Russell Sage Foundation.

Duncan, Greg J. and Aaron J. Sojourner. (2012). "Can Intensive Early Childhood Intervention Programs Eliminate Income-Based Cognitive and Achievement Gaps?"

European Commission. (2014). *Study on the effective use of early childhood education and care in preventing early school leaving*, Publications Office of the European Union, Luxembourg.

European Commission. (2014). "Proposal for key principles of a Quality Framework for Early Childhood Education and Care", *Report of the Working Group on Early Childhood Education and Care*.

Executive Office of the President of the United States. (2014). *The Economics of Early Childhood Investments*.

Felfe, C. and Lalive, R. (2011). *How Does Early Childcare Affect Child Development? Learning from the Children of German Unification*. CESifo Area Conference on Economics of Education: Center for Economics Studies.

Friedman, S.L., Scholnick, E.K., Bender, R.H., Vandergrift, N., Spieker, S., Hirsh Pasek, K., Keating, D.P., Park, Y. and the NICHD Early Child Care Research Network (2014), Planning in Middle Childhood: Early Predictors and Later Outcomes. *Child Development*, 85

Garces, Eliana, Duncan, Thomas, and Janet Currie. (2002). "Longer Term Effects of Head Start". *American Economic Review* 92(4).

Gialamas, A., Mittinty, M.N., Sawyer, M.G., Zubrick, S.R., Lynch, J. (2014). Child Care Quality and Children's Cognitive and Socio-emotional Development: an Australian Longitudinal Study. *Early Child Development and Care* 184.

Haskins, R., and Barnett, S. (2010). *Investing in Young Children: New Directions in Federal Preschool and Early Childhood Policy*. National Institute for Early Education Research.

Havnes, T., & Mogstad, M. (2009). No child Left Behind – Universal Child Care and Children's Long-Run Outcomes, *IZA Discussion Paper* No. 4561.

Heckman, J.J., Moon, S.H., Pinto, R. and Savelyev, P.A. (2009) 'The rate of return to the HighScope Perry Preschool Program', *The Journal of Public Economics*, vol 94, no 1-2.

Heckman, J.J., Pinto, R., and Savelyev, P.A. (2012). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes (No. w18581). *National Bureau of Economic Research*.

Heckman, James J., Seong Hyeok Moon, Rodrigo Pinto, Peter A. Savelyev, and Adam Yavitz. (2010). "Reanalysis of the High/Scope Perry Preschool Program." *Quantitative economics*, 1(1).

Heckman, J. J. (2006). Skill formation and the Economics of Investing in Disadvantaged Children, *Science* 312 (5782).

Kautz, T., Heckman, J.J., Diris, R., ter Weel, B. and Borghans, L. (2015). *Fostering and Measuring Skills: Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success*. OECD Publications, Paris.

Lazzari A. & Vandenbroeck, M. (2012). “Literature review on the participation of disadvantaged children and families in ECEC services in Europe”, in Bennet: *Early childhood education and care (ECEC) for children from disadvantaged backgrounds: Findings from a European literature review and two case studies*, DG EAC.

Mashburn, A., Pianta, R., Hamre, B., Downer, J., Barbarin, O., Bryant, D., Burchinal, M., Early, D. and Howes, C. (2008) “Measures of Classroom Quality in Prekindergarten and Children’s Development of Academic Language, and Social Skills”. *Child Development* 79(3).

Morris, A. (2011) *Student Standardised Testing: Current Practices in OECD Countries and a Literature Review*, OECD Education Working Paper No. 65, Paris.

Naudeau, S., Naoko, K., Valeria, A. and Neuman, M. (2011). *Investing in Young Children: An Early Childhood development Guide for Policy Dialogue and Project Preparation*. The International Bank for Reconstruction and Development/The World Bank.

Organisation for Economic Co-operation and Development (OECD) (forthcoming). *Monitoring the Quality of ECEC Systems*.

Organisation for Economic Co-operation and Development (OECD) (2015). *Skills for Social Progress: The Power of Social and Emotional Skills*, OECD Publishing, Paris.

Organisation for Economic Co-operation and Development (OECD) (2014). *Education at a Glance 2014: OECD Indicators*, OECD Publishing, Paris.

Organisation for Economic Co-operation and Development (OECD) (2013). The Draft Proposal on the Data Development Strategy of the OECD Network on Early Childhood Education and Care, 13th Meeting of the OECD Network on Early Childhood Education and Care (ECEC), 17-18 June 2013, Paris.

Organisation for Economic Co-operation and Development (OECD) (2012). *Starting Strong III: A Quality Toolbox for Early Childhood Education and Care*, OECD Publishing, Paris.

Organisation for Economic Co-operation and Development (OECD) (2011). *PISA in Focus: Does Participation in Pre-primary Education translate into Better Learning Outcomes at School?*, OCED Publishing, Paris.

Organisation for Economic Co-operation and Development (OECD) (2006). *Starting Strong II*, OECD Publishing, Paris.

Puma, M., Bell, S., Cook, R., & Heid, C. (2012). *Third Grade Follow-up to the Head Start Impact Study-Final report*, (OPRE Report). Office of Planning, Research and Evaluation Administration for Children and Families U.S. Department of Health and Human Services.

Ramey, Garey & Ramey, Valerie A. (2010). “The Rug Rat Race”. Brookings Papers on Economic Activity 41, no. 1.

Romano, E., Babchishin, L., Pagani, L.S., and Kohen, D. (2010). School readiness and later achievement: Replication and extension using a nationwide Canadian survey. Developmental Psychology, 46(5).

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I. and Taggart, B. (2012) *Effective Pre-school, Primary and Secondary Education 3-14 Project (EPPSE 3-14): Final Report from the Key Stage 3 Phase: Influences on Students’ Development from Age 11-14*, Research Report DfE-RR202, London: Department for Education.

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I. and Taggart, B. (2008) *Final Report from the Primary Phase: Preschool, School and Family Influences on Children's Development during Key Stage 2 (7-11)*, Research Report DCSF-RR061, London: Department for Children, Schools and Families.

Sylva, K., Melhuish, E., Sammons, P., Blatchford, I.-S., & Taggart, B. (2004). *The Effective Provision of Pre-School Education [EPPE] Project* (Project Final Report).

Timperley, H. (2008). Teaching: Professional Learning and Development, *International Academy of Education*.

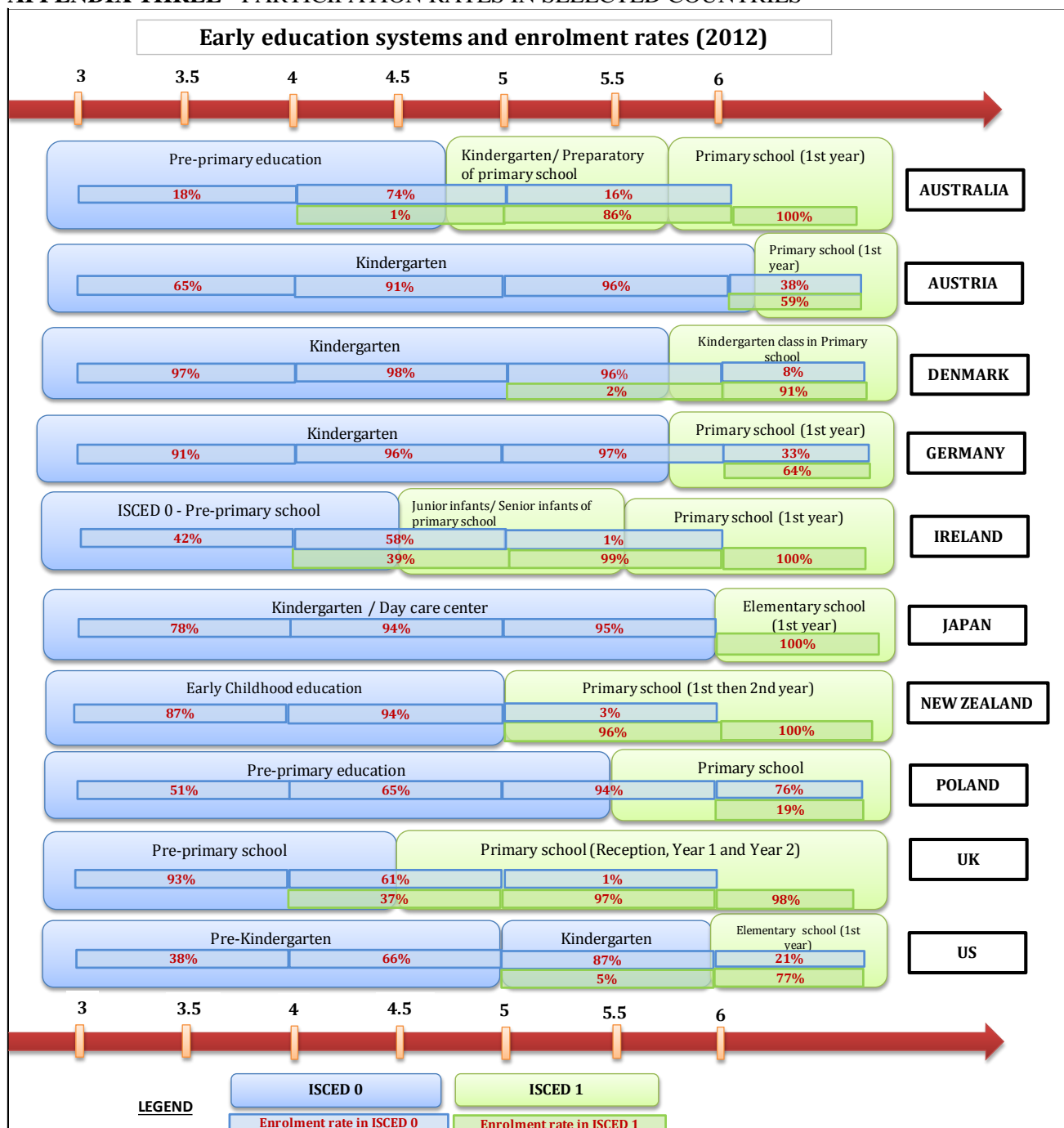
Vandell, D.L., Woolfe, B. (2000). *Childcare Quality: Does It Matter and Does It Need to Be Improved?*

Institute for Research and Poverty, Madison, Wisconsin.

Walker, D., Greenwood, C., Hart, B., and Carta, J. (1994). Prediction of school outcomes based on early language production and socioeconomic factors. *Child development*, 65(2).

Walters, Christopher. (2014). *Inputs in the Production of Early Childhood Human Capital: Evidence from Head Start*. Working Paper 20639. Cambridge, Mass.: National Bureau of Economic Research

APPENDIX THREE - PARTICIPATION RATES IN SELECTED COUNTRIES



Source: Table C2.1 in Education at a Glance (OECD, 2014)

Notes: The diagram does not necessarily reflect the compulsory starting age in a certain stage, but rather the common practices in terms of starting age.

• Australia: The age entry requirements for the last year of preschool and for the first year of school vary across jurisdiction over a 6 months period.

• UK: ISCED 0 widely differs across UK nations. In Wales, early education is provided in the "Fondation Phase", which spans 3 to 7 years-olds children and which covers pre-school as well as the first 3 years of primary school (Reception, Year 1, Year 2). Most of 3 and 4 years olds are in maintained schools. In England, the Early Years Foundation Stage (EYFS) sets standards for development, learning and care of children from birth to the age of 5. Most children are outside maintained schools.

MINIMUM GENERAL CONDITIONS FOR OECD CONTRACTS

The following articles constitute of the minimum general conditions of the contract to be signed between the OECD and the Contractor to whom the Call for Tenders would have been awarded (the "Contract"). These minimum general conditions are not exclusive and could, as the case may be, be modified and/or complemented with additional conditions in the Contract.

ARTICLE 1 – GOODS OR SERVICES

The goods and/or services provided under the Contract (hereinafter "The Work") shall strictly comply with the standards mentioned in the Terms of Reference. It is expressly agreed that the Contractor shall perform the Work in strict accordance with all Standards or, where no such standards have yet been formulated, the authoritative standards of the profession will be the applicable norms.

ARTICLE 2 - PRICES

Prices charged by the Contractor for the Work shall not vary from the prices quoted by the Contractor in its Tender, with the exception of any price adjustment authorised in the Contract.

ARTICLE 3 - PAYMENTS AND TAXES

Payment will be made in Euros.

In case the Contractor is located outside of France, the Organisation is exempt from taxation, including from sales tax and value added tax (V.A.T.). Therefore, the Contractor shall not charge any such tax to the Organisation. All other taxes of any nature whatsoever are the responsibility of the Contractor.

ARTICLE 4 - DELAY IN EXECUTION

The Contractor shall perform the Work in accordance with the time schedule and the terms specified in the Contract, this being an essential element of the Contract. Any delay will entitle the Organisation to claim the payment of penalties as negotiated between the Contractor and the Organisation.

ARTICLE 5 - ACCESS TO THE PREMISES

If the Work requires at any time the presence of the Contractor and/or of the Contractor's employees, agents or representatives ("Personnel") on the premises of the Organisation, they shall observe all applicable rules of the Organisation, in particular security rules, which the Organisation may enforce by taking any measures that it considers necessary.

ARTICLE 6 - IMPLEMENTATION OF THE WORK

The Contractor undertakes that the Work shall be performed by the individual(s) named in the Contract or otherwise agreed in writing by the Organisation. The Contractor may not replace said individual(s) by others, without the prior written consent of the Organisation.

MINIMUM GENERAL CONDITIONS FOR OECD CONTRACTS

ARTICLE 7 - AUTHORITY

The Contractor hereby declares having all rights and full authority to enter into the Contract and to be in possession of all licences, permits and property rights, in particular intellectual property rights, necessary for the performance of the Contract.

ARTICLE 8 - LIABILITY

The Contractor shall be solely liable for and shall indemnify, defend and hold the Organisation and its personnel harmless from and against any and all claims, losses, damages, costs or liabilities of any nature whatsoever, including those of third parties and Contractor's Personnel, arising directly or indirectly out of or in connection with Contractor's performance or breach of the Contract.

It is the responsibility of the Contractor to possess adequate insurances to cover such risks, including any risks related to the execution of the Contract.

ARTICLE 9 - REPRESENTATIVES

Neither the Contractor nor any of its Personnel:

- shall in any capacity be considered as members of the staff, employees or representatives of the Organisation;
- shall have any power to commit the Organisation in respect of any obligation or expenditure whatsoever;
- shall have any claim to any advantage, payment, reimbursement, exemption or service not stipulated in the Contract. In particular and without limitation, it is understood that neither the Contractor, nor any of the Contractor's Personnel may in any manner claim the benefit of the privileges and immunities enjoyed by the Organisation or by its personnel.

ARTICLE 10 - INTELLECTUAL PROPERTY

The copyright and any other intellectual property rights arising from the Work carried out in performance of this Contract, including the intermediate and final results thereof, shall, on an exclusive and worldwide basis, automatically vest in the Organisation as the Work is created, or be assigned to the Organisation, as the case may be under any applicable legal theory. The price agreed between the Contractor and the Organisation is deemed to include this transfer of rights.

The Contractor undertakes not to use the Work for any purpose whatsoever that is not directly necessary to the performance of the Contract, except with the prior written consent of the Organisation. The Contractor shall ensure that the Contractor's Personnel are expressly bound by and respect the provisions of the present clause.

ARTICLE 11 - TRANSFER OF RIGHTS OR OBLIGATIONS

MINIMUM GENERAL CONDITIONS FOR OECD CONTRACTS

The Contractor shall not transfer to any third party any rights or obligations under this Contract, in whole or in part, or sub-contract any part of the Work, except with the prior written consent of the Organisation.

ARTICLE 12 - TERMINATION

Without prejudice to any other remedy for breach of Contract the Organisation may claim, the Organisation reserves the right to terminate the Contract without any prior notice or indemnity:

- i) in the event of failure by the Contractor to comply with any of its obligations under the Contract; and/or
- ii) if the Contractor, in the judgment of the Organisation, has engaged in corrupt or fraudulent practices in competing for or in executing the Contract.

The Organisation may also, by written notice sent through registered mail with recorded delivery to the Contractor, terminate the Contract, in whole or in part, at any time for its convenience. The notice shall specify that termination is for the Organisation's convenience, the extent to which Work of the Contractor under the Contract has been completed, and the date upon which such termination becomes effective. The Work that is complete on receipt of notice by the Contractor shall be accepted by the Organisation, at the Contract terms and prices. For the remaining, the Organisation may elect:

- i) To have any portion completed at the Contract terms and prices; and/or;
- ii) To cancel the remainder and pay to the Contractor the amount corresponding to the completed work.

ARTICLE 13 – FINANCIAL INFORMATION

During the Contract and at least seven years after its termination, the Contractor shall :

- i). keep financial accounting documents concerning the Contrat and the Work ;
- ii). make available to the Organisation or any other entity designated by the Organisation, upon request, all relevant financial information, including statements of accounts concerning the Contrat and the Work, whether they are executed by the Contractor or by its any of its subcontractors.

The Organisation or any other entity designated by the Organisation may undertake, including on the spot, checks related to the Contrat and/or the Work.

ARTICLE 14 - ARBITRATION CLAUSE

Given the status of the Organisation as an international organisation, the rights and obligations of the Contractor and the Organisation shall be governed exclusively by the terms and conditions of the Contract.

OPTION 1: In case there is no transfer of goods, services or funds outside of the OECD host country (France):

MINIMUM GENERAL CONDITIONS FOR OECD CONTRACTS

Any dispute, controversy or claim arising out of or relating to the interpretation, application or performance of this Contract, including its existence, validity or termination, shall be settled by final and binding arbitration in accordance with the Permanent Court of Arbitration Optional Rules for Arbitration between International Organizations and Private Parties, as in effect on the date of this Contract. The number of arbitrators shall be one. The language to be used in the arbitral proceedings shall be English. The place of arbitration shall be Paris (France). The appointing authority shall be the Secretary-General of the Permanent Court of Arbitration. The Parties expressly agree that if an award rendered by the arbitral tribunal is annulled, the jurisdiction ruling on the annulment proceedings cannot rule on the merits of the case. The dispute will therefore be settled by new arbitral proceedings in accordance with this clause.

OPTION 2: In case of transfer of goods, services or funds outside of OECD host country (France):

Any dispute, controversy or claim arising out of or relating to the interpretation, application or performance of this Contract, including its existence, validity or termination, shall be settled by final and binding arbitration in accordance with the Permanent Court of Arbitration Optional Rules for Arbitration between International Organizations and Private Parties, as in effect on the date of this Contract. The number of arbitrators shall be one. The language to be used in the arbitral proceedings shall be English. The place of arbitration shall be Paris (France). The appointing authority shall be the Secretary-General of the Permanent Court of Arbitration. The Parties expressly renounce their right to seek the annulment or setting-aside of any award rendered by the arbitral tribunal.

Nothing in the Contract shall be construed as a waiver of the Organisation's immunities and privileges as an international organisation.

ARTICLE 15 – CONFIDENTIALITY

Any information, on any medium whatsoever, sent to the Contractor to which the Contractor obtains access on account of the Contract, shall be held confidential. In consequence, the Contractor shall not disclose such information without the written prior consent of the Organisation. The Contractor shall ensure that the Contractor's Personnel is expressly bound by and respect the provisions of the present clause.

ARTICLE 16 - DURATION OF THE CONTRACT

Unless otherwise stated in the Call For Tenders, the duration of the Contract shall be for one year. It may be renewed twice by tacit agreement for periods of one year, but the total duration may not exceed three years.