Network on Early Childhood Education and Care

INDICATORS OF LEARNING AND WELL-BEING ENVIRONMENTS FOR CHILDREN

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

This paper has been prepared by the ECEC Network’s Thematic Working Group on Data, which is led by Mr. Richard Walley (New Zealand) and comprised of the following participants: Ms. Anne-Marie Dieu (Belgium-Fédération Wallonie-Bruxelles), Ms. Christele Van Nieuwenhuyzen (Belgium-Flanders), Mr. Bernhard Kalicki and Mr. Philipp-Laurens Rogge (Germany), Mr. László Limbacher (Hungary), Ms. Tove-Mogstad Slinde (Norway) and Ms. Rosalyn Harper (United Kingdom), for the 12th meeting of the ECEC Network held from 17-18 December 2012 at OECD headquarters in Paris.
INDICATORS OF LEARNING AND WELL-BEING ENVIRONMENTS FOR CHILDREN

Discussion Paper Presented to the 12th Meeting of the OECD Network on ECEC

Summary

This paper is intended to support discussion at the 12th meeting of the OECD Network on ECEC on indicators of learning and well-being environments for children. This contributes to strand 2 of the network’s work programme.

It explores four key areas:

1. Why are learning and well-being environments for children important? What information is important for policy formation?
2. The purpose and use of trans-national indicators.
3. Principles that could be applied to the development of indicators.
4. Discussion of further work on possible indicators – what do we have now, and what will be useful in future?

To progress strand 2 of the network’s work programme, the 12th meeting needs to discuss and agree:

1. Which current indicators are useful and require quality improvement?
2. What new information is required to fill any gaps?

This paper is complemented by another paper exploring existing data and indicators in more detail.

1. Learning and well-being environments for children

This paper examines a broad range of topics based around the concept of learning and well-being environments for children. For the purposes of this paper, we treat this topic broadly. Rather than focus on a definition of the overall heading, the concept is used to capture a wide range of information and ideas, and applied to those things we would consider to be ‘quality’ factors, both process and structural. We refer in particular to variables around staffing and the physical learning environment.

This wide range of information is nonetheless central to policy development in ECEC. It is these factors that translate initial Government action – access and expenditure – into the expected outcomes. Whether the anticipated outcomes are basic wellbeing (health, safety, security), individual development, contemporaneous wellbeing and quality of experience, school preparedness, or any other of the range of possible goals set out in policy and curriculum statements, the mechanisms by which they are delivered are similar.

Across these learning and wellbeing items, the policy challenges faced by most jurisdictions are similar. ECEC is a developing field, still lacking in many commonly accepted policy positions supported by robust evidence. So, first and foremost, any new information must address at least some of these policy challenges. These can be summarised as research questions as follows–

- What works? What aspects of learning and wellbeing environments genuinely improve the learning and wellbeing of young children?
• **What works best?** Are the greatest improvements in learning and wellbeing environments to be had from improving qualifications, numbers of staff, ongoing professional development, or any other factors?

• **How do Governments make it work?** What is the mixture of expenditure, regulation, and information across various domains of activity that achieve meaningful gains? And at what cost?

• **How are the gains distributed?** In the field of ECEC, one of the key accepted facts is that gains should accrue particularly to children from disadvantaged backgrounds. How effective are different systems at achieving the kinds of distribution of resources, learning and wellbeing opportunities that should achieve these gains.

An understanding of the comparative impact of these mechanisms is therefore extremely important for policy formation. With a greater understanding of the translational mechanisms that comprise learning and wellbeing environments, we can begin to understand the landscape of key variables and areas for future focus, and in turn design policies that have a greater impact, potentially at a lower cost.

Finally, as practitioners in an under-explored area of public policy, it is fair to say that we don’t know what we don’t know. The potential for extended and useful information to emerge that highlights new possible avenues of enquiry is significant.

2. **Purpose and Use - why are we focused on indicators?**

We are not short of indicators in ECEC. In fact, significant time and effort, often overlapping, is periodically invested in investigations of the ‘state’ of ECEC in various jurisdictions, and often trans-nationally.

Recently reports, statistics and indicators have been published in EAG, by the Economist research unit, and as part of the UNICEF Innocenti benchmarking process, while nationally a number of sophisticated systems exist for monitoring and comparison.

Our first point for development of indicators on quality of learning environments should therefore be – what can the OECD add to this picture? The key value propositions would seem to be:

• Organisational infrastructure that supports gathering, analysing and publishing education and care data.

• A wide spread of jurisdictions involved for comparison.

• A broad base of support, and ability to produce ‘definitive’ information and methodologies, as well as reports on relevant themes for quality and output of education, for example, from EAG and TALIS.

• The ability to derive analysis from a broader base of statistical information (e.g. PISA, economic data, labour market data, other trans-national indicators of progress and wellbeing).

• General face validity and reputation – the OECD can be seen in some areas to be the most reliable source of information and analysis.

**Trans-national indicators**

There is a clear difference between the utility and nature of indicators developed for purposes of domestic policy development, and trans-national indicators. The former are often developed to monitor a particular policy initiative, or are designed around inherent variability in domestic policy settings. For example, New Zealand measures the proportion of qualified, registered teachers in teacher-led ECE
services as a) it has adopted a domestic goal of 80% of staff being registered and qualified in 2012, and b) there is variability in performance on this measure and it changes over time (some services have more teachers, some less).

Trans-national indicators, particularly in education, tend to have different functions. These can be summarised as follows; note, this is a broad-brush characterisation and not a detailed analysis.

1. They tend to serve to monitor change over time only in the broadest sense. Jurisdictions will usually know their domestic results long before they are compared to other jurisdictions, and will often be some way into policy initiatives in response to the information before the trans-national results are known. Only a few trans-national education indicators (the key example being PISA, and a range of indicators present in EAG) function as measures of domestic progress, and it is characteristic of these that they are administrated and collected centrally rather than reported domestically.

2. Their principle function is comparison between jurisdictions, and they act as starting points for policy analysts to undertake deeper analysis of the underpinning factors that lead to differential performance.

3. As such, they tend to function often as formative research inputs, rather than summative outcome indicators. They are used to answer research questions such as ‘why has jurisdiction X done so well on Y’, ‘how has jurisdiction Z achieved A’, or simply ‘how are we doing in comparison?’.

We note this because it is important to consider the eventual use to which indicators will be put. Our interest in the ECEC policy network as OECD members is primarily the domestic utility we can derive from its activities.

As such, a primary recommendation of this paper is that our overarching goal should be to seek indicators for, in this case, learning and well-being environments, that:

- enable genuine, useful comparison between jurisdictions, and
- facilitate formative research that enhances our individual and collective knowledge of best practice in ECEC policy development.

It also seems obvious, but is nonetheless worth stating, that the utility of trans-national indicators is enhanced in line with the number of jurisdictions able to provide high-quality data regularly. As noted above, this should in theory be an advantage of the OECD as a focal point for collection, but one that should be considered carefully at the 12th network meeting, as not all jurisdictions will be able to report in detail on all suggested indicators.

3. Suggested principles for development of indicators

Data collection is necessarily a reductive process (and monitoring – for the purposes of this paper, the repeated collection of similar data over time). That is to say, it cannot represent the entirety of any particular state or condition. Instead, it consists of a part of what is measured. The hope is that this part provides a reliable enough representation of an aspect of the whole to enable evaluative, or at least comparative judgements to be made.

It is for this reason that we often talk about ‘indicators’ – they are facts intended to indicate the existence of a broader state than is represented integrally by the facts presented.

This makes choice of which data to collect important. For two reasons. First, collection of unrepresentative data obviously affects the fundamental utility of the process as a whole. Useful
conclusions cannot be drawn from unrepresentative data. Second, it risks formation of poor policy. Although initially we strive to measure what is important, inevitably, and quickly, what we measure becomes important. The ongoing debates about the utility of GDP as a measure of economic success are one example of a place where the indicator drives policy formation, to a far greater extent than the obverse.

From this discussion we may distil at least one principle – that of **representative validity**. To what extent does a set of data represent the concepts which interest us?

As we are more specifically considering comparative monitoring, a further important aspect of representative validity is one of reliability over time and range. A strong indicator will represent a change in state over time or between jurisdictions whilst remaining reliable. We may add to this that it should be equally representative at all measured points, although this can, to a certain extent, be compensated for by choice of representative media and some statistical controls.

To this, we can add a second principle, one of **representative utility**. There is difficult territory to negotiate here between two further important concepts; implied normative judgements, and the relationship between those judgements and actual effect.

For example, if a rate is used as an indicator, in the absence of other information it is easy to assume that the limit of the range (e.g. 100% or 0%) is a normative optimum, and jurisdictions should be considering policies to reach the implied goal. Even without rates, comparative quantitative indicators tend to imply a normative judgement without further contextualisation, usually either ‘lower is better’ or ‘higher is better’. Where this is the case, it is obviously of paramount importance that these things are true.

**Representative validity explored through ratios**

Representative validity, over time and across the range of the scale, bears some further examination through an example. A common area for focus in ECEC indicators is the ratio of adults to children. This is currently reported in Education at a Glance, under indicator C2.3 (OECD, 2012). Measurement in 2010 puts the range of this indicator across the OECD at between a little over 6, and a little over 25.
When considering ratios, lower numbers are assumed to be good, underpinned by evidence suggesting this is a strong indicator of the quality of educational experiences. Broadly, this gives it high representative validity over an end state. One would expect, ceterus parabus, children in systems with lower numbers are experiencing better learning and wellbeing environments (or having better experiences of those environments) than those in systems with higher numbers.

From the working group, individual jurisdictions treat this indicator in different ways. New Zealand relies more heavily on regulated minimum ratios for purposes of policy formation than the reported average from EAG. Similarly, the federal government and scientists in Germany focus much of their policy or evaluative attention on differences in regulated ratios between Länder. In contrast, the child care sector in Flanders does not only consider ratios as a headline indicator of learning or wellbeing environments (it is however set as a condition to be able to provide quality), preferring a provider-administered evaluative instrument.

Domestic policy considerations aside, it is useful to consider the utility of the EAG measure of ratios. Validity over time and range is potentially highly questionable, because the nature of the movements along this scale will make significantly variable differences to education experience depending on starting point. A move from 24 to 23 will make little difference to education experiences (on average a 4% change to staff face time per child), whereas a move from 5 to 4 will make a significant, 20% difference to staff face time. In short, one would expect a significantly accelerated change in the thing represented towards the bottom of the scale, but zero change towards the top. The indicator stops being functional in measuring meaningful change over time or comparisons between jurisdictions beyond a certain point (we could conjecture maybe an adult:child ratio of 15:1, after which point the change in a child’s learning or wellbeing caused by the change in the indicator will asymptotically approach zero as ratios increase).

It is reasonable to note at this point with this example that these problems are not insurmountable. Representative validity could be improved by condensing the asymptotic part of the scale (e.g. adopting a ‘15+’ category), or by picking a measure that works better across a linear scale – percentage of staff time per child, for example.

A second point to consider is the use of an average ratio. A mean necessarily smooths differences across the range. If we are interested in disparity within jurisdictions, as well as between, as we often tend to be, a single national mean will be functionally only indicative – even a representation of the true situation will require further investigation and interpretation. For example, some fundamental questions that could be asked around a mean adult:child ratio are:

- How many children experience this ratio? How many experience higher or lower ratios? What is the standard deviation?
- What are the characteristics of children that experience ratios deviating from the mean? Are they wealthier, poorer, from particular ethnic or immigrant groups?
- What are the characteristics of settings offering different ratios? Are there difference ratios for different age groups, different ratios for different types of ECEC provision?
- Where jurisdictions have regulated for ratios, how do the reported ratios deviate from the regulated? To what extent does practice reflect the intent of regulatory or funding regimes?

**Representative utility explored through ratios**

Exploring again the current EAG measure of ratios, it is easy to spot potential pitfalls. First, the normative assumption is that lower ratios are good. This may well lead an abstract viewer to conclude that the end point of the range – a ratio of one-to-one (or lower!) – is the natural end state policy goal.
But we may very sensibly question whether this is the case. We conjecture above that, for example, a move from five to four will confer significant benefits for the learning and wellbeing environment. Will a move from two to one confer the same benefits? This is a different point to the point made above about linear reliability across a scale. The key question here is whether a normative goal is implied, and whether that goal is genuinely optimal. This example is a solidly debatable, although highly abstract, point; practically, it is hard to conceive of any jurisdiction implementing widespread policies around such a ratio.

Nonetheless, the overall question remains important. Many members of the network have as their main business advising Ministers or politicians. Ministers have a strong tendency to look at a table of comparative statistics and assume either that they should be near the top (or bottom), or ask a more sensible question of advisers – “where should we be?”. Absent a specific domestic policy goal, this can be challenging to answer if the indicator implies an end state not necessarily related to actual desired outcomes.

4. What should we measure?

The OECD revised project plan proposes a set of specific areas/indicators as a programme of work on data development of the ECEC Network [EDU/EDPC/ECEC(2012)3/REV1]. The proposal was developed, drawing upon the range of measures current in use [EDU/EDPC/ECEC(2012)3/REV1/ANN2] as well as the recommendations set out in this paper.

Having considered some objectives and principles for indicators, we are now in a position to consider what may be useful to measure and compare. A classic framework for considering objectives and points of relevance is that of inputs, outputs, and outcomes. Simply, this can be characterised as – we spend money on a thing – that thing produces something – something happens as a result.

ECEC policy development is characterised by a focus on inputs. This is partially attributable to the formative stage of policy development of many jurisdictions; it is natural to focus attention on inputs, as the subsequent stages simply aren’t possible without them. The classic ‘iron triangle’ of quality in ECEC – group size, ratios, and qualifications – is entirely input focussed.

ECEC also has an existing useful construct for delineating inputs and outputs; the concepts are directly analogous to structural and process quality. Outcomes have tended to be a more difficult area to address in ECEC, and have historically been considered in three ways; as specified in curriculum documents, as measured by later individual characteristics (academic achievement, health, wellbeing), and as considered through the framework of a child’s contemporary experience and immediate wellbeing in an ECEC setting.

A summary of the kinds of areas that may be explored under the three headings is noted below, with some notes where an indicator, or some data, already exists. This is a broad summary of the area by generalities, and is not intended as a proposed or exhaustive list of possible indicators.
We noted at the start of the paper that one of the principle items of utility for international indicators is as a formative input into domestic policy analysis formation. The structure noted here provides some strong guidance as to how this might be implemented – via a set of indicators that enables us to explore questions about the links between inputs, outputs, and outcomes.

Although already covered in the OECD paper, some features are notable from this summary.

1. There is a very obvious bias towards input indicators. Output and outcome indicators are highly underdeveloped.

2. Much input information is to be found in Starting Strong 3 (OECD, 2012), and is hence not subject to regular collection by the OECD or other organisation (although Euridyce holds some richer data for European jurisdictions, e.g. (European Commission, 2009)).

<table>
<thead>
<tr>
<th>Items that could be measured</th>
<th>Existing indicators or information?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Inputs</strong></td>
<td></td>
</tr>
<tr>
<td>Staff</td>
<td>Pay</td>
</tr>
<tr>
<td></td>
<td>Conditions</td>
</tr>
<tr>
<td></td>
<td>Qualification</td>
</tr>
<tr>
<td></td>
<td>Turnover</td>
</tr>
<tr>
<td></td>
<td>Professional registration</td>
</tr>
<tr>
<td></td>
<td>Staff competence</td>
</tr>
<tr>
<td></td>
<td>Staff well-being</td>
</tr>
<tr>
<td>Pay ranges and some qualification information reported in Starting Strong 3. Education at a Glance 2012 also includes some data on teacher wages in ECEC.</td>
<td></td>
</tr>
<tr>
<td>Qualifications</td>
<td># and level</td>
</tr>
<tr>
<td></td>
<td>Mixture within a setting</td>
</tr>
<tr>
<td>Reported in Starting Strong 3. Eurydice also collected some data on staff qualifications.</td>
<td></td>
</tr>
<tr>
<td>Adult:child ratios</td>
<td>Mean reported in EAG 2012, while regulated ratios are included in Starting Strong 3. Eurydice also has data available on ratios.</td>
</tr>
<tr>
<td>Group size</td>
<td>Mean reported in Starting Strong 3</td>
</tr>
<tr>
<td></td>
<td>Some data available within Eurydice.</td>
</tr>
<tr>
<td>Physical Environment</td>
<td>Space per child</td>
</tr>
<tr>
<td></td>
<td>Access to resources (toys, books, play equipment &amp;c)</td>
</tr>
<tr>
<td></td>
<td>Access to trips, visits</td>
</tr>
<tr>
<td>Mean reported in Starting Strong 3</td>
<td></td>
</tr>
<tr>
<td>Availability, quality of curriculum</td>
<td>Some indicative aspects reported in Starting Strong 3 (subject area focus). Eurydice also has some information available on content of curricula.</td>
</tr>
<tr>
<td>Availability, structures around provision and delivery of professional development</td>
<td>Some data published in Starting Strong 3.</td>
</tr>
<tr>
<td>Expenditure</td>
<td>Covered in some detail in EAG 2012 and earlier versions, including by age, public/private, per capita and a number of other measures. Family database and Social Expenditure database also have a variety of data on ECEC financing, expenditures and funding of ECEC.</td>
</tr>
<tr>
<td>Outputs</td>
<td>Quality of pedagogy (in systems with an explicit education focus)</td>
</tr>
<tr>
<td></td>
<td>Quality of attachments, interactions</td>
</tr>
<tr>
<td></td>
<td>Quality of child experience</td>
</tr>
<tr>
<td>Outputs</td>
<td>Recorded in PISA age 15</td>
</tr>
<tr>
<td></td>
<td>Child wellbeing / involvement</td>
</tr>
<tr>
<td></td>
<td>Contemporaneous outcomes</td>
</tr>
</tbody>
</table>
The range of input information so far collected is heavily focused on the workforce, with significant gaps around group size and environment.

Although not explored explicitly in this paper, a common feature of data collections is an emphasis on explicitly ‘educational’ programmes (pre-school instruction), and less on care, or education, for younger children, or on integrated programmes. EAG actually goes to some lengths to exclude both younger children and family day care arrangements from its indicators.

These features of currently available data present significant challenges across the objectives and principles of the exercise; extracting reliable, useful, representative data from which to make formative analytical judgements will be challenging.

In fact, it may well be the case that without an initial spread of indicators across these measures, we will be poorly placed to make informed decisions on representative validity and utility. It would appear that the current range of input-focused indicators does not serve us well in terms of making judgements on what is actually important, what normative benchmarks we are safe to imply, and where the key thresholds are for calibration of quantitative indicators (if indeed such thresholds can be said to exist).

A case study on the way one jurisdiction has considered these challenges is presented below.

### Case study – Flanders

A useful case study and point of analysis is the way Flanders has developed instruments to provide information across measures of inputs, outputs, and outcomes in 0-3 child care settings.

To assist judgements about whether a service should be accredited, the independent care inspectorate administers an instrument, *KWAPOI*, to judge the quality of a facility. This consists of three parts.

- **A checklist** is used to consider input factors; administration, staffing, staff-child ratio, accommodation, safety, hygiene, feeding and medical treatment of the children.

- Output measures are obtained through **observation scales**. Inspectors use these to observe how all the staff in the facility treat the children (and where appropriate the parents) when the children are welcomed on arrival, during meals, when they are giving physical care and doing toilet-training, during rest periods, and during activities and play. Inspectors score staff for each of these 5 situations. They base their scoring on the seven elements that are important for educational interaction between care workers and children: flexibility, individualisation (adapting to the child), stimulation, structuring (offering security), promoting independence, offering freedom of action and safety).

- Output quality is further considered through a **semi-structured interview**, to give the facility an indication of areas that still need to be improved.

Crossing the boundary between outputs and outcomes is a different, non-compulsory instrument, SICS, a Process-oriented Self-evaluation Instrument for Care Settings. SICs considers how much the children feel at home in their child care facility (well-being) and how much they enjoy the activities and are actively absorbed in them (involvement). SICS enables service providers to improve their approach in specific ways.

Finally, an individual outcomes-based instrument, SICS –Cf, enables facilities to follow up an **individual child**. Where SICS helps facilities to evaluate their general approach, this instrument allows the staff to evaluate an individual child in order to develop an approach specific for that child.

For the future, Flanders plans to develop an entirely new composite instrument to consider the pedagogical quality of child care settings.
Items for discussion

To continue with strand 2 of the network’s work programme, at the 12th meeting of the network, we must answer two key questions on indicators of learning and wellbeing environments for children. These are:

1. Which current indicators are useful and require quality improvement?
2. What new information is required to fill any gaps?

Which current indicators are useful and require quality improvement?

While the range of available data, either gathered through EAG, Starting Strong 3, or other research sources is wide, much of it could potentially fail to provide useful information for the purposes of formative research or further policy analysis. The chief area for improvement of these data is a cohesive picture of the impact of the policy decisions that have led to the situation reported, and the subsequent impact of that situation on children, their learning, and wellbeing.

Key weaknesses around current collections appear to be the ability to differentiate across experience within jurisdictions, and the applicability of a solid evidence base to inform best practice, and hence range or calibration of indicators.

For example, considering again the example of adult – child ratios. In New Zealand, ratios in most centre-based settings are regulated at 10:1 for children two and over, and 5:1 for children under two. For the purposes of Education at a Glance, New Zealand reports a ratio of 7.14, produced by dividing the number enrolments by the number of staff for children over three. So, the 5:1 ratio for younger children is ignored in this indicator, as is the interesting fact that most de facto ratios exceed regulated ratios by some margin.

This experience is mirrored in Belgium. Education at a Glance reports a ratio of 16:1, the ratio for pre-school children. The ratio for children under 2½ is lower, and also changes across types of service (1:5 for childminders and 1:7 for childcare services). Some information also suggests that de facto ratios may be higher than regulated ratios in pre-school classes.

Of strong further interest, in addition to this more finely graduated information, would be, for example, the qualifications of these staff (are services offering better ratios at the expense of qualifications?), and in particular distribution by child characteristics. Is a child more likely to experience these above-average ratios if they are from a higher-income household? Are parents paying higher fees in services with better ratios?

Such information would begin to inform some of the most basic questions raised at the start of this paper –

- How do Governments make it work? What is the mixture of expenditure, regulation, and information across various domains of activity that have achieved this situation? And at what cost, to who?
- How are the gains distributed? How effective are different systems at achieving the kinds of distribution of resources, learning and wellbeing opportunities that should achieve strong gains for children?
So the first recommendation for discussion is that existing input indicators be improved with the addition of common differentiation factors. The selection of key indicators is relatively obvious and already present. It could consist of:

- Levels of staff qualification
- Mixture of qualification levels within settings
- Group size
- Adult:child ratios
- Space per child

Most of which is available either in EAG or SS3. However, the utility of these indicators would be greatly improved by differentiation factors covering how these indicators appear different in response to various items. A fuller delineation could read as follows:

1. Levels of staff qualification
   a. ISCED levels
   b. Minimum (regulated) standard and de facto standard
   c. Variation by child age
   d. Variation by course (qualification) content
   e. Variation by type of setting (family day care, kindergarten, child care, etc)
   f. Variation of experience by family income
   g. Experience of children with special education needs
   h. Variation of experience by ethnicity / immigrant / refugee status

2. Mixture of qualification within settings
   i. ISCED levels
   j. Minimum (regulated) standard and de facto standard
   k. Variation by child age
   l. Variation by course (qualification) content
   m. Variation by type of setting (family day care, kindergarten, child care, etc)
   n. Variation of experience by family income
   o. Experience of children with special education needs
   p. Variation of experience by ethnicity / immigrant / refugee status

3. Group size
   q. Minimum (regulated) standard and de facto standard
   r. Variation by child age
   s. Variation by type of setting (family day care, kindergarten, child care, etc)
   t. Variation of experience by family income
   u. Inclusion of children with special education needs
   v. Variation of experience by ethnicity / immigrant / refugee status

4. Adult:child ratios
   w. Minimum (regulated) standard and de facto standard
   x. Variation by child age
   y. Variation by type of setting (family day care, kindergarten, child care, etc)
   z. Variation of experience by family income
   aa. Variation of experience by ethnicity / immigrant / refugee status

5. Space per child
   bb. Minimum (regulated) standard and de facto standard
   cc. Indoor and outdoor
Variation by child age
Variation by type of setting (family day care, kindergarten, child care, etc)
Variation of experience by family income
Variation of experience by ethnicity / immigrant / refugee status

A key limiting factor here would be the availability of these data domestically. Formative work undertaken by this working group suggests that domestic data collection is potentially not well-developed, and therefore reporting on these differentiation factors (and indeed some of the indicators themselves) may be challenging.

It is therefore useful that network participants come prepared to discuss, on behalf of their jurisdictions:

- Whether these data are available now
- Whether they are useful
- Whether the jurisdiction would be prepared to improve (or consider improving) domestic data collection to the state that they will be able to report on these factors in the medium term (three-five years).

Further metrics for learning environments

One factor missing from this list which is important is other environmental quality factors. A number of reliable quality instruments, notably ECERS-R (Harms, Clifford, & Cryer, Early Childhood Environmental Rating Scale, Revised Edition (ECERS-R), 1998), carry scales that measure the quality of the static environment – furniture, toys, physical learning opportunities - in more detail than a measure of space available. A reasonable amount of research suggests that these items, particularly availability of books and other literacy aides, can be important inputs into children’s experience.

Measuring these aspects of learning environments can be particularly important for policy formation for three reasons. Firstly, they are items on which information for policy makers is often scarce, although a number of jurisdictions regulate for these items in a variety of ways. Secondly, availability of a wide range of environments and learning resources is linked to learning and wellbeing; these are important aspects of the experience of children. Finally, for children from disadvantaged backgrounds, it is often access to an environment strongly geared towards learning and wellbeing that makes a significant difference. Simply, if your home has few books, you are less likely to develop precursor literacy skills and strong foundation linguistic competencies.

Items in this category measured by ECERS-R are a good starting point for considering further input metrics. These are:

2. Furniture for routine care, play and learning
3. Furnishings for relaxation and comfort
4. Room arrangement for play
5. Space for privacy
7. Space for gross motor play
8. Gross motor equipment
15. Books and pictures
In common with the range of indicators noted above, it would be useful to be able to delineate further the range of indicators above against:

a. minimum (regulated) and de facto standard
b. child age
c. variation by type of setting
d. Variation of experience by family income
e. Variation of experience by ethnicity / immigrant / refugee status

A second recommendation for discussion is the development of separate, or composite measures for quality of environment based on indicators like those described above. In particular, how well-placed will jurisdictions be to contribute data to such an indicator?

**What new information is required to fill any gaps?**

This is a significant question for further discussion. A further conclusion of this paper is that the scarcity of output and outcome measures across our available indicators limits the utility of all indicators for policy formation purposes.

Drawing further on the questions about policy utility posed at the start of this paper, consideration of outputs would bring us much closer to the ability to answer some of the most important policy questions in the field of ECEC. To re-state, these are:

- **What works?** What aspects of learning and wellbeing environments genuinely improve the learning and wellbeing of young children?
- **What works best?** Are the greatest improvements in learning and wellbeing environments to be had from improving qualifications, numbers of staff, ongoing professional development, or any other factors?
- **How do Governments make it work?** What is the mixture of expenditure, regulation, and information across various domains of activity that achieve meaningful gains? And at what cost?
- **How are the gains distributed?** How effective are different systems at achieving the kinds of distribution of resources, learning and wellbeing opportunities that should achieve these gains?

As noted, the ability to link input measures with output measures would represent a significant step forward in our ability to make and evaluate effective ECEC policy.

**Output indicators**

Indicators around outputs would potentially require the development and implementation of a common collection instrument. The current lack of such information is as much to do with practical problems as anything else. Some one-off studies link across input-focused instruments (such as ECERS-R) and output-focused instruments (such as CLASS or CIS) (Tietze, Cryer, Bairrao, Palcios, & Wetzel, 1996), and in addition we note successive versions of and addenda to ECERS-R in particular focus more on output as well as input measures (e.g. staff and adult activity have some items, particularly in ECERS-E) (Siraj-Blatchford, Sylva, & Taggart, 2006). But the intensive nature of observation required to produce results from these instruments could make their use for collection of regular indicators impractical.
This raises a number of obvious implementation challenges; who would administer, how often, which instrument to use, &c and so forth. These are key points for discussion for the entire network. These discussions could be guided by the following questions.

- **Can we commission research into domestically-collected indicators of output quality?** In particular, what are the possibilities for comparative research into the validity and reliability of instruments to do so, particularly international (such as CLASS) and domestic (such as IQS, an integrated quality scale developed in Germany consisting of some adaptations of ECERS-R and – E with the addition of specific novel items).

- **How may jurisdictions have evaluative functions able to provide overview data?**

- **How would we calibrate the data from such functions?** Is it worth considering commissioning an exercise, e.g. from the OECD secretariat, to do so?

There are some existing models of ways a selection of output indicators could be collected from participant jurisdictions. A sampling technique could be adopted (jurisdictions do not need to provide data on all services), and jurisdictions with existing evaluative functions, such as New Zealand and the UK, may find it more straightforward to incorporate such data collection into their existing activities. Nonetheless, even for jurisdictions with well-developed evaluative functions that cover a reasonable proportion of the services delivered regularly, this could represent a large shift in the way their activities are conducted.

A shorter-form variant of this could be a study similar to TALIS conducted in the ECEC sectors of member countries. This approach has the advantage of being relatively low-impact and easy to administer; however, as a self-administered survey of professionals, there are also inherent data collection limitations. A further interesting approach could be the collection of reported wellbeing data from children themselves; again, a sampling and survey approach could be usefully employed.

Possible exemplar indicators in this context are challenging to derive sensibly. This is principally because the instruments from which they would be derived (e.g. ECERS variants, CLASS, and CIS) are well-tested composite instruments with a strong supporting evidence base. Removing specific items from subscales and considering them in isolation as indicators would almost certainly do much to degrade their representative validity and utility.

Rather, it is simplest at this stage to consider the possibility of a trans-national indicator that consists of a reasonably frequently measured composite indicator. For example, jurisdiction mean and distributed scores could be given for ECERS overall, and separate indicators provided for subscales, e.g.

**Standard**

- Space and furnishings
- Personal Care Routines
- Language-Reasoning
- Activities
- Interactions
- Program Structure
- Parents and Staff
Curriculum

- Reading
- Mathematics
- Science and Environment
- Diversity

In addition, and in line with the suggested direction noted above, all of these factors could then be reported on as varied by child age, type of setting, family income, ethnicity / immigrant / refugee status, etc.

A similar set out of output indicators could be considered along the same lines, for example the subscales from CIS –

- Sensitivity
- Harshness
- Detachment
- Permissiveness

And again, varied by pertinent factors within jurisdictions.

We wish to emphasise that these items are not a definitive list of suggested indicators. While tools like ECERS and CIS are predominant in the field, and it makes sense to start with those to provide a sense of what could be possible, the more important conclusion here is that a regular or semi-regular trans-national study into outputs using well-tested instruments will be required to produce the kinds of indicators that will be useful for ECEC policy development. The challenge is not in identifying desirable indicators, but in identifying the desirable process and infrastructure to collect the information from which the indicators would be derived.

Outcome indicators

The working group has had some initial discussions about the possibility and nature of outcome indicators across ECEC. These discussions have reached two summary conclusions, as this is obviously a matter for much further discussion and careful design work. These are:

1. The construction of outcome indicators is broadly desirable for the purposes of policy utility. They could provide strong data to enable genuine, useful answers to be developed to the key policy utility questions posed early in this paper – in particular, what works, and what works best. However, the development of such indicators must be approached with extreme care; a narrow summative assessment focus is unlikely to yield useful results.

2. In particular, due attention should be given to the very broad range of outcomes many jurisdictions expect their ECEC systems to contribute to.

On this latter point, applying the concept of representative validity makes development of indicators of outcomes particularly challenging. This is because, for an indicator to be representative of a particular outcome, it is important first to define the desired outcome.

ECEC systems tend not have homogenous outcomes across jurisdictions. Broadly, they could be considered across a number of different categories. These can include, but are not limited to –

- precursor and early literacy and numeracy skills
• individual wellbeing
• learning dispositions and skills
• social, cognitive, behavioural and physical development
• school preparedness
• curriculum subject areas (Starting Strong 3 lists 16, e.g. figure 2.5)
• broader health and developmental outcomes
• specific intervention or learning for particular groups, e.g. language learning for indigenous or immigrant populations, intervention for children with special education needs
• the level of satisfaction of parents and staff with care, learning and wellbeing environments can be considered important outcomes in their own right; and indeed, these may be more straightforward to gather data on than more complex individual outcomes.

It is important to note that, for many of these outcomes, measurement, if possible, tends to be complex, technical and intensive. It is for this reason that the contributors to this paper caution specifically against narrow, summative approaches to outcomes; the risk will be high that such approaches will cover only a small proportion of the desired outcomes of an ECEC system, and not with a high degree of accuracy.

To return to delineation of these content areas, they also reside within three or four significant expected outcomes from ECEC. These could broadly be described as –

• Intrinsic future (e.g. improved learning or wellbeing for children after they have left ECEC – part of which PISA measures now)
• Intrinsic contemporaneous (e.g. children enjoy positive learning and wellbeing during their attendance at ECEC)
• External future (e.g. reduced overall rates of crime, health, social problems)
• External contemporaneous (e.g. essentially labour market participation, or allowing parents the opportunity to undertake activities other than childcare)

In fact, Governments generally expect all four of these outcomes, in various mixtures, from their ECEC systems, although the emphasis may change from time to time. The representative validity of an indicator measuring future externalities from a system designed substantively to facilitate parent labour market participation could easily be criticised for not measuring the thing the system was designed to achieve.

In addition, a definitive description of these anticipated outcomes can also be challenging for jurisdictions to compile. Expectations are often set out in a wide range of policy documents, curricula, strategies and other such outputs, and can in addition sometimes be defined far more at a local level than central Government.

This framework also presents two key challenges. Measuring external outcomes presents causal challenges (are changes in indicators a genuine measure of activity in ECEC, or the product of some other factor), and measuring future outcomes present both causal and timing challenges (e.g. PISA 2012 will report on the impact of ECEC as delivered in and around 2002 for most jurisdictions). All of these matters can seriously affect the representative validity of any outcome indicators proposed.
Finally, it is worthwhile considering again the key policy utility questions posed earlier in this paper. These were:

- **What works?**
- **What works best?**
- **How do Governments make it work?**
- **How are the gains distributed?**

As noted above, some notion of the outcomes achieved is vital for considering questions of what works, and what works best. But a full and thorough discussion of a common meaning for these questions is necessary. We will find it challenging to determine what works if we do not first have a common understanding of what ‘works’ actually means; and the brief summary above suggests that this may yet be very different for different jurisdictions.

**Summary / Conclusion**

This paper has provided a number of key questions for discussion at the 12th network meeting, and some initial guidance and frameworks for considering how to approach these questions. In sum, consideration should be given to –

- The improvement of existing input indicators by the addition of differentiation factors (and potential addition of further input indicators).
- The development of indicators of outputs – almost certainly to be achieved through the implementation of a trans-national study of some description.

A further item for discussion should be prioritisation of the potential recommendations arising from the meeting. While some jurisdictions may consider many of the items suggested desirable, practicalities will inevitably mean that some may be progressed quickly, and some less so. Even the addition of differentiation factors to existing indicators may challenge the data collection capabilities of some jurisdictions, and may have to proceed slowly and additively.

**Contributors to the Paper**

Anne-Marie Dieu  
Bernhard Kalicki  
Christele Van Nieuwenhuyzen  
Laszlo Limbacher  
Philipp-Laurenz Rogge  
Richard Walley  
Rosalyn Harper  
Tove-Mogstad Slinde
EDU/EDPC/ECEC(2012)4

BIBLIOGRAPHY


