

**The Early Years Enriched Curriculum  
Evaluation Project:  
Rationale for the Research Design**

Sproule, L., Trew, K., Rafferty, H., Walsh, G.\*,  
O'Neill, B., M<sup>c</sup>Guinness, C. & Sheehy, N.

**School of Psychology,  
Queen's University Belfast &  
\*Stranmillis University College**

**March 2003**

## 1. Context

The Early Years Enriched Curriculum commenced with multiple, inter-related objectives. These objectives expanded as interest in the curriculum gained momentum. In 2001 there were six schools and nine classes within the Belfast Education and Library Board working with the enriched curriculum. By February 2003 there were approximately 80 schools involved with many others wanting to join.

The pace of expansion raised issues, such as the consistency and integrity of the programme across Education and Library Boards, not foreseen when the curriculum was launched. The first end of year evaluation report addressed these considerations and included several recommendations that constituted part of the formative evaluation for the Northern Ireland CCEA. That feedback has had an important role in the rapidly changing situation: the formative aspect of the evaluation has become part of a sophisticated model for change in early years education in Northern Ireland in which each of the stakeholder groups have an significant role. The model of change incorporates many features of well-established good practice in educational innovations and is discussed more fully in Appendix A.

In the context of the objectives for the enriched curriculum the evaluation methodology incorporates several approaches to address the four objectives specified in the CCEA tender:

1. Assess the short and longer-term impact of the Enriched Curriculum on the learning dispositions and progress in literacy and numeracy of the initial cohort of pupils in the six schools in the Greater Shankill area through Years 2, 3 and 4, as compared with an appropriate control group. (Note: Year 1 was covered in the first stage of the project in 2000 – 01.)
2. Assess the short and longer-term impact of the Enriched Curriculum on the learning dispositions and progress in literacy and numeracy of a smaller cohort of Year 1 pupils drawn from contrasting areas of Northern Ireland through Years 2 and 3 as compared with an appropriate control group
3. Investigate the perceptions of teachers and parents regarding the appropriateness of the revised curriculum.
4. Determine some of the factors that impact on the effectiveness of curriculum implementation, for example, teacher in-service training, external support and resourcing, and examine how these factors influence the implementation of the revised curriculum.

The cohort of schools referred to under Objective 2 were selected by CCEA from the group that adopted the Enriched Curriculum in its second year. It was intended that they should be broadly representative of schools in all the Education and Library Boards (ELBs). This group of schools constitutes a second study within the overall evaluation. It is referred to as the Contrasting Areas (CA) group of schools to distinguish it from the first group. The characteristics of this group will be described in greater detail later on.

Good evaluations of educational innovations do not limit their focus to outcome measures. Outcome measures provide the clearest evidence of an effect but they provide little or no information on the manner of implementation. The major focus of the evaluation of the Enriched Curriculum is on the quantification of outcomes. However, the evaluation is designed to provide the CCEA with detailed knowledge of the strengths and weaknesses of the programme *as it is being implemented*.

## 1.1. Design considerations

The identification of appropriate control groups was identified as a problem at an early stage. The problem was caused by: (i) the leakage of aspects of the Enriched Curriculum to potential control-group schools; and (ii) the fact that the social conditions of schools using the enriched curriculum could not be exactly matched with those of control group of schools<sup>1</sup>.

To address this problem it was decided to use year-ahead and two year-ahead pupils as controls throughout the second phase of the implementation of the enriched curriculum. In this scenario, each class experiencing the Enriched Curriculum is matched with *very similar* classes: those that are one and two years ahead in the same school, who received the pre-existing curriculum. Use of these controls minimises the influences of potential confounding effects, such as differences in ethos between schools, and may permit some estimate of teacher effects, provided teachers continue teaching the same year group throughout.

In addition to these in-house controls, the use of standardised measures means that Enriched Curriculum groups can be systematically compared with a very large representative national sample from the UK and with specific sub-groups such as schools in disadvantaged inner-city areas.

The other major strands of the evaluation are concerned with recording the views of teachers and parents. As the Enriched Curriculum was being rolled-out it became clear that it had implications for all the teachers in Key Stage 1. Thus, considerable importance is attached to assessing the views of each year group of teachers as they encounter the Enriched Curriculum for the first time. The continuity of the research team over the life of the Enriched Curriculum has deepened our understanding of the nature of the Enriched Curriculum and the issues it raises. This has enabled us to refine the teacher interview schedules each year and to probe significant issues in depth. As a result, the evaluation team has input to the literacy working group, which is presently preparing a written framework for the Enriched Curriculum together with suitable guidance for teachers.

Parent's views on the Enriched Curriculum are fundamentally important and these are being gathered through interviews and questionnaires in both sets of schools. These parents have proven to be very insightful about the experience of their children. In order

---

<sup>1</sup> For a complete discussion, see Sproule, Rafferty, Trew, Sheehy, McGuinness and Walsh, 2001.

to encourage parents to report their concerns and any negative reactions to the Enriched Curriculum the evaluation team has provided guarantees of anonymity as well as reassurances that their views will be recorded and made publicly available.

As the Enriched Curriculum moved into its second phase a further important development was the creation of a Communications Committee. The Committee was formed to facilitate good relations between all the professional parties involved. The role of the communications committee is discussed in more detail in Appendix A.

## **1.2. Instruments**

The selection of measurement instruments was made against four criteria:

- (i) Functionality: Will the instrument provide the type of information needed to answer the research question?
- (ii) Quality: Is the instrument valid and reliable?
- (ii) Developmental suitability: Is the instrument appropriate to the developmental levels of the school children and the practical requirements of the classroom?
- (iii) Cost effectiveness: Does the measurement instrument represent good value for money?

A variety of tests and instruments were reviewed against these criteria. Those that were selected are being used to address Objectives 1 and 2 of the evaluation. At the core is a robust measure of basic attainment, the Performance Indicators in Primary Schools (PIPS). This is augmented by a range of age-appropriate measures assessing learning disposition, language and reading experience.

### **1.2.1 Selection of PIPS**

The Enriched Curriculum is innovative and there are no measures of attainment that precisely match its content and outcomes. Thus, it was decided to:

1. Baseline the children using well-established, robust tests that samples many known predictors of later achievement.
2. Track the children using tests that are appropriate for the pre-existing curriculum until such time as the old and Enriched curricula merge again.
3. Supplement these well-established tests with instruments that measure the particular goals of the Enriched Curriculum.

The PIPS suite of age-appropriate tests was chosen as a cost-effective measure with high reliability and validity. PIPS measures achievements across a wide range of domains pertinent to the National Curriculum. The PIPS database was first established in 1993 and now contains data on hundreds of thousands of children in the UK in more than 4000 schools. This means that the achievements of children who receive the Enriched

Curriculum can be compared against: (i) the national average and (ii) schools that are comparable on a range of socio-economic indicators.

At baseline, PIPS tests the child in:

- Skill at writing his or her own name
- Picture vocabulary (often used by educational psychologists as a proxy IQ measure)
- Ideas about reading (e.g. reading moves from top to bottom of the page)
- Ability to select rhyming words from a given selection (phonological awareness)
- Ability to repeat sounds, including nonsense sounds
- Knowledge of letters
- Word recognition skills
- Knowledge of basic mathematics concepts
- Counting
- Cardinality and invariance aspects of number
- Subtraction with picture cues
- Addition with picture cues
- Arabic numeral recognition
- Two and three-digit number recognition

Thus, the test covers a wide range of known predictors of later achievement, accounting for about half the variance in scores at the end of Year 2<sup>2</sup>. The composite score across all domains in the test, the PIPS standardised total score at baseline, correlates significantly with achievement of children following the National Curriculum in England. Nationally, the correlation with mathematics achievement at the end of Year 2 (Northern Ireland's Year 3) is .65 and with reading achievement is .70 (Tymms, Merrell & Henderson 1998).

In the second year, besides picture vocabulary, reading and mathematics attainment tests, PIPS includes a test of non-verbal ability. This test, taken in conjunction with the test of picture vocabulary, generates a 'context score' that can be regarded as a proxy IQ score. This score provides useful background against which to gauge the progress of each child and will be used to assess whether specific sub-groups of children do well under the Enriched Curriculum.

In the second year of PIPS the reading scores attained are partially abstracted from a task which requires comprehension. In later years, reading for understanding dominates the reading tasks.

---

<sup>2</sup> A child can complete the baseline PIPS assessment in about twenty minutes on average, thus minimising disruption for the teacher and child.

## 1.2.2 Implementation of PIPS

In the context of the mismatch between PIPS, which is calibrated to the objectives of the National Curriculum, and the innovative content of the Enriched Curriculum, PIPS scores at the end of the first and second years provide a pointer to the extent to which the implementation of the Enriched Curriculum has actually taken place. Teachers who adhered very strictly to the Enriched Curriculum could expect *lower* scores on average than those who reverted to the traditional curriculum. A *drop* in standardised scores would also suggest that parents are not circumventing the new curriculum, for example by teaching the children to read at home using techniques associated with the traditional curriculum.

PIPS generates two value-added scores for each child during each year. The value added scores represent the degree of change compared with the child's previous performance and compared with his or her proxy IQ scores.

With regard to the control groups, the choice of the year-ahead and two-year-ahead option was made possible by the availability of PIPS baseline data on children in the Shankill schools from the earlier Greater Shankill Study. We do not have data for these children in their second year but comparison will be possible in the third and fourth years and for the two-year-ahead group only, at the end of the first year.

Testing of the children is undertaken by trained testers who are either psychologists or teachers and who are extensively briefed on administration of the tests. The baseline tests have shown an impressive consistency in baseline PIPS totals across the three consecutive year groups.

The situation is not quite so straightforward in the second group of schools – the Contrasting Areas schools. Firstly, baseline PIPS data are not available on control group children (although one school has shared PIPS data gathered in previous years by the teachers). Secondly, there is considerable variation in the entry profiles from school to school. For example, it is clear from the baseline scores of the Enriched Curriculum cohort in these schools, that they all differ considerably from the Shankill schools, having noticeably higher baseline totals on average. Thirdly, the different Education and Library Boards have somewhat different approaches to the Enriched Curriculum. The implication of the last two points is that it may be necessary to treat each of the schools as a case study: This means that it would not be permissible to amalgamate the data from all schools in a statistically valid manner, (something which can be done with schools in the Shankill study). Nevertheless, the evaluation design will allow statistically valid conclusions to be drawn about the performance of children undertaking the Enriched Curriculum. This can be done in four ways:

- Comparison of second and third year PIPS scores of Enriched Curriculum children and controls (and of performance on all the other tests)
- Comparison of each child's PIPS scores and proxy IQ scores

- Comparison of each individual's progress with national average progress of children with the same baseline score.
- Comparison of each individual's progress with national average progress of children in schools with a similar catchment area.

In addition, a small study of the stability of baseline scores in one of these schools is under way but not all the data will have gathered until near the end of the project. If the scores turn out to be stable across the lifetime of the evaluation study, it will enable us to make further inferences.

### **1.2.3 Supplementary attainment tests**

A prominent goal of the Enriched Curriculum is the improvement of oral language skills but PIPS does not address oral language directly. Accordingly, a supplementary test of oral language was sought. The test selected was the Bus Story Test, which is part of the Renfrew Language scales (Renfrew Language Scales 1969). In this test the researcher tells the child a story with picture prompts and then the child retells the story while following the same pictures in the stimulus booklet. Three scores are abstracted from the recording of the child's story: an information score reflecting the content, the average length of the five longest sentences and the number of subordinate clauses used. Thus, the test not only addresses the maturity of the child's expressive language skills, but also the child's understanding of the story and ability to remember it. These are all aspects of language which are important for progress in reading with understanding. In Year 3, it is expected that children may begin to find the Bus Story Test somewhat childish. It may therefore not be appropriate to continue using it.

At baseline and the end of the first year, the section devoted explicitly to mathematical concepts in PIPS is somewhat short. The Enriched Curriculum eschews recorded arithmetic and concentrates on the development of a thorough understanding of basic math concepts. Accordingly, in the first and second years, it was decided to augment the PIPS test programme with the age-appropriate form of the Boehm Test of Basic Concepts. This assesses the understanding of many of the concepts needed for early mathematical development.

At the end of Key Stage 1, it is expected that the Enriched Curriculum will prove to have been a vehicle for improvement in creative writing skills. This would be a consequence of three factors. First, teachers already believe there is greater appreciation of and wider experience of books. Second, this enhanced experience of books should interact with children's developing self-confidence in narrative writing, achieved through proper valuation of the child's early efforts. Third, improvements in oral language skills are predicted to augment performance in creative writing. Thus, at this stage, it is appropriate to include a test of creative writing skills. The WOLD (Wechsler Objective Language Dimensions 1996) Test of Written Expression was selected as a suitable robust measure. It is only age appropriate from the age of eight years.

The WOLD test of written expression will be administered at the end of Year 4 in Shankill schools, and possibly at the end of Year 3 in Contrasting Area schools if teachers believe the children can attempt it. It is a test of creative writing on a given topic. The test generates scores on all of the following scales:

- Ideas and development
- Organisation, unity and coherence
- Vocabulary
- Sentence structure and variety
- Grammar and usage
- Capitalisation and Punctuation

Spelling is not a focus of the test. It is intended that children will not have their creative ability quenched by concerns about correct spelling and this is explained to them in the instructions for the test. In addressing creative writing, the WOLD addresses a domain not covered by the PIPS test. In addition, it addresses the following aspects of learning in the Enriched Curriculum:

- The ability to tell a coherent story
- The influence of exposure to a wide variety of printed material, including fiction
- The ability to adhere to instructions within a written task
- Increasing sophistication of oral language (since this is a prerequisite for good written language)
- The conventions of written as opposed to oral language
- The acquisition of vocabulary

#### **1.2.4 Attitudes and self-esteem**

The Enriched Curriculum makes a point of fostering both the child's self-esteem, in relation to learning and in general, and positive attitudes to all aspects of work. It is well established that positive attitudes to reading are necessary for optimum progress in reading. The Reading Self-Concept Test (Fredrickson and Cameron 1995) looks at three aspects of attitude; the child's perceived self-competence in reading, perception of the difficulty of reading work and enjoyment of reading.

The Harter Test of Self Competence is a well-established measure of a child's self-esteem in four areas; with his or her peers, with the mother (or significant other), in schoolwork and in physical play.

Teachers are also asked to rate children on six scales, one each for effort in work, confidence in work, relationship with peers, general happiness and adjustment in school, attention skills and behaviour problems. Some of these provide a crosscheck for the child's self-rating in the tests above.

## 1.2.5 Other pertinent data gathered on the children

### *Social and Behavioural measures*

During the first year of the project, teachers were asked to complete the Adaptive Social Behavioural Inventory (the ASBI scale, Hogan, Scott and Bauer, 1992) for each child. This scale is in use in the Effective Provision of Pre-School Education Project (EPPE), a major government-funded 6-year longitudinal study which is currently looking at pre-school provision in England. A Northern Ireland equivalent, the EPPNI study, also uses this scale. Completion of this scale constituted a significant burden for teachers at a time when they were enduring considerable additional workloads. The research co-ordinator leading the evaluation at that time was only able to get the data on 59 children back from teachers and that was achieved only after exerting persistent pressure on them to complete the task. Such a situation does not make for those good working relationships between the researchers and teachers that are essential for a well-executed, rigorous evaluation. Accordingly, it was agreed by the Communications Committee that teachers should be given a less burdensome alternative. They rated each child against a four-point scale on the following items:

- Makes an effort with his or her work
- Is confident about his or her work
- Gets on well with other children
- Seems happy and well adjusted
- Has problems maintaining attention
- Has behaviour problems

Each of these items has clear face validity in relation to the stated goals of the Enriched Curriculum. When all the data have been collected, it may be possible to perform factor analysis on the data. This, in turn, may allow us to assign a composite score<sup>3</sup> to each child, representing that child's adaptability to the teaching experience. A preliminary analysis of the data available to date has indicated some external validation for the composite score. There is a statistically significant association between the composite scores and scores on the ASBI scale ( $r = .65, p < .01$ ). Finally, the composite adaptability score was a statistically significant predictor for the reading and mathematics scores at the end of Year 2. The correlations are summarised in Table 1.

---

<sup>3</sup> A two-factor model, one factor related to social skills and one related to attitudes to work is also a possibility. Only factor analysis of the complete data set will reveal which model is a better fit for the data.

*Table 1*  
*ASBI and Adaptability Social and Behavioural ratings: Correlations with outcome measures for Enriched Curriculum children at the end of Year 2*

	‘Adaptability’	End Year 2 maths	End Year 2 reading	ASBI
‘Adaptability’		.318*	.424*	.635*
End Year 2 maths	.318*		.338*	.157*
End Year 2 reading	.424*	.338*		.366*
ASBI	.635	.157*	.366*	

\*Significant at the  $p < .01$  level

*Other data gathered*

The following data were also gathered for each child:

- Special needs status
- Attendance
- Free school meals status
- Percentage of free school meals in the whole school<sup>4</sup>
- Socio-economic status for the child’s postcode, or for the school if the former was unavailable
- Teacher-child ratio in each year of the project

### **1.3 Addressing Objective 4: Factors impacting the curriculum**

This objective is being addressed by documentation of all resources, training schedules and external support networks. A review of the research literature indicated that there were no off-the-shelf measures available to quantify these influences. However, it was clear that a process measure would be a component of the evaluation, particularly in light of the need to determine whether classroom change had actually taken place and if so, what kind of effects it might have on the child’s experience. The evaluation rationale was to use a measure of classroom process in combination with a study in which a sample of children experiencing the Enriched Curriculum are systematically compared with a sample of controls. The preferred measure was Walsh’s QLI, in its original form and in the form adapted for other year groups. The research design for this objective is presented in Table 2.

---

<sup>4</sup> This has been shown to be a significant predictor of outcome measures in the Croxford and Cowie (1999) study of Aberdeen schools.

*Table 2*  
*Research design for Objective 4*

<b>School Year</b>	<b>Autumn</b>	<b>Spring</b>
2001-02	Documentation of resources, training and external support.	Walsh's QLI for Primary 1 classes Walsh's adapted QLI for Primary 2 classes.
2002-03	Documentation of resources, training and external support.	Walsh's QLI for Primary 1 classes Walsh's adapted QLI for Primary 2 classes. Walsh's adapted QLI for Primary 3 classes
2003-04	Documentation of resources, training and external support.	Walsh's QLI for Primary 1 classes Walsh's adapted QLI for Primary 2 classes.

## **2. Observational measures of process: General issues**

### **2.1 Functionality**

Observational measures are most useful for providing rich data on the process of educational interventions. Questionnaires and interviews can deliver information on respondents' perceptions of what is happening in the classroom, although these are liable to be distorted by misunderstandings, expectations, wishful thinking or similar influences. Further, the researcher may have failed to predict all of the ways in which the intervention may succeed or may fail.

A criticism often levelled against qualitative measures relates to the impossibility of conducting 'blind' observational studies in educational interventions. An observer who is sufficiently qualified to be an expert judge will almost always notice that an educational intervention is taking place. The observer may then become subject to bias. The size of the bias can be reduced by using several observers.

### **2.2 What types of information will observational measures provide?**

Observational measures make a large contribution to the formative aspects of evaluation. Chiefly, they supply useful information in three areas. Firstly, they provide information about the actual delivery of the intervention. Theoretically, it would be possible for a teacher to feel committed to a new programme and fail to change her delivery in any major way. In practice, the delivery of a new curriculum is determined by a teacher's personal views, their professional experience and expertise and their reception of the training on a new curriculum. Teachers differ and they can be expected to differ in the way they deliver the new curriculum. What are these differences? Only an observational measure of classroom process can detect these.

Children differ too and they can be expected to differ in the way they respond to the new curriculum. Measures of outcome for children are what matter for assessing achievement but they yield relatively little information on the learning environment of children. Observational studies of classrooms can address this limitation.

Finally, the observational measures provide triangulation for the subjective reports from parents and teachers.

### **2.3 Validity and reliability**

To be useful, an observational instrument must be valid, reliable, applicable to the particular situation in question and cost-effective.

The validity of a test represents the extent to which it is a sound measure of whatever quality or qualities it is supposed to measure. Content validity refers to whether the measure samples those behaviours which are important in determining success in the individual/s who are observed.

Reliability is related to the stability (also known as intra-observer reliability or consistency), reproducibility (also known as inter-observer reliability) and accuracy of the recording instrument. It is usual to undertake a calibration study to determine reliability.

Reliability does not guarantee validity and there is often a tension between the two requirements.

### **2.4 What observational instruments are available?**

The only instrument which was designed specifically to look at early years primary school classrooms is the Quality of Learning Instrument (QLI), designed by Dr Glenda Walsh of Stranmillis University College. A paper fully describing the development process may be found in Appendix B. The evaluation team also considered the closest alternative prospect, the ECERS-R (Harms, Clifford and Cryer 1998) and its adapted form the ECERS-E (Sylva, Siraj-Blatchford and Taggart 1999). These have been similarly designed to look at classroom process, but specifically for classes of nursery school age. As one would expect, there are some strands common to the ECERS and QLI instruments but the QLI is vastly superior in its match to the early years situation in primary schools. Further, the QLI is unique in recognising that several aspects of learning, namely the teacher's performance, the children's responses and the learning environment<sup>5</sup>, are determinants of the quality of the child's experience. The nearest alternative measure, the ECERS scales, are more restricted in scope and focus on the

---

<sup>5</sup> Environment here implies more than the physical environment. It encompasses the learning atmosphere and the use of resources. There is a dynamic interaction between the three aspects of the child's experience.

nature of the physical classroom environment and neglect a serious consideration of the curriculum in action.

## **2.5. What counts as good value for money?**

In determining a budget allocation within a given project, there is always a tension between the quantitative and qualitative aspects of the evaluation. The aim is to prioritise the elements of the evaluation in relation to the wishes of the funding body (given that these are consistent with the high quality of the evaluation). Various factors contributed to the cost of carrying out the work:

- Cost of any materials
- Cost of training and employing skilled staff
- Length of the observation
- Cost of analysis

Approximately 85% of financial resources are directed to the quantitative assessment of outcomes and 15% to observational assessments: of the children's learning environment, parents perceptions and teachers perceptions (see section 3.5 below for further details).

## **3. Observational measures: Specific issues**

### **3.1 How may observational measures address Objective 4?**

The ECERS scales were the only potentially suitable alternatives to Walsh's QLI and these were assessed and rejected against the functionality criterion: they do not provide the type of information needed to answer the research question.

### **3.2 What types of information do observational measures provide?**

Walsh's QLI was originally specifically designed to address quality in Year 1 classrooms. It was later adapted for Year 2 and Year 3 classrooms. The ECERS-R and the ECERS-E were designed for nursery education settings, and their content reflects the nursery curriculum rather than a Key Stage 1 curriculum.

In addition, the QLI directly addresses many of the goals of the Enriched Curriculum:

- The development in the children of motivation to learn (good learning dispositions)
- The development of the children's powers of concentration
- Development of appropriate independence within the learning process
- Development of children's confidence in their own ability to succeed in learning
- The children's acquisition of social skills

- Development of children’s respect for people and property
- Development of children’s social and emotional character

The QLI also addresses indirectly the Enriched Curriculum’s concentration on giving children the tools to succeed in education rather than the traditional curriculum’s concentration on passive knowledge acquisition. This aspect is addressed through the following two scales:

- Higher-order thinking skills (HOTs)
- Multiple skill acquisition

Within the QLI, the rater – the person completing the observational analysis - does not merely assign a number to each of the 27 scales of the instrument. They also make notes of observations in support of each decision. These observations contain valuable information on the classroom process. In addition they provide evidence to support the rater’s decision.

### 3.3 How valid are the available measures?

Walsh’s QLI has been validated by internationally recognised early years experts (See Appendix B).

### 3.4 How reliable are the available measures?

A calibration study was carried out on Walsh’s QLI as part of its development as a measure suitable for classroom observation. The study, based on ten experts of education in the early years, is described in the paper in Appendix B. It reported very high levels of inter-rater reliability but did not include a computation of a single measure of inter-rater reliability between all those who participated. Neither did it specifically address the competence in the use of the instrument in respect of the researchers taking part in the Enriched Curriculum project. Accordingly, a further study was carried out to quantify the inter-rater reliability within this group and to compare their implementation of the scoring performance with that of the early years experts. For the purpose of this study, all twenty of the Enriched Curriculum researchers watched a training video and allocated scores for each of the 27 measures in Walsh’s QLI. For each variable, alpha, the coefficient of inter-rater reliability described by Krippendorff (1980) was computed. These coefficients are presented in Table 3. The formula for alpha is given by:

$$\alpha = 1 - \sum_{\substack{\text{all pairs} \\ \text{raters}}} \left( \frac{\text{extent of observed mismatch}}{\text{mismatch expected due to chance}} \right)$$

*Table 3*  
*Alpha (coefficient of inter-rater reliability) for all elements of the QLI*  
*Data from twenty EYECEP researchers*

	Children's actions	Teaching strategies	Environment
Motivation	.80	.77	.69
Concentration	.80	.84	.69
Confidence	.77	.69	.77
Independence	.77	.77	.71
Well-being	.77	.68	.58
HOTs*	.69	.84	.60
Multiple skills	.84	.69	.74
Social interaction	.89	.69	.60
Respect	.80	.84	.77

\* Higher order thinking skills

Unlike some inter-rater reliability coefficients, Krippendorff's alpha takes account of not only the correlations between raters' scores, but also of the preponderance of each grade within the data matrix<sup>6</sup>. This is equivalent to taking into account the degree of severity of each rater's set of scores and the degree to which the raters make use of the whole scale.

Table 3 shows that there are very high levels of agreement among the observers in respect of the children's actions and the teaching strategies. No value of alpha drops below .69 for these scales. The somewhat lower levels of agreement in respect of the environment relate to aspects of well-being, higher-order thinking skills and social interaction and are caused by the limitations imposed by using the video presentation. It was not possible to judge these aspects of the environment from the video as accurately as one would have wished. Dr Walsh presented some written information to compensate for this deficiency of data but this written information would necessarily be less rich than personal experience within the classroom.

In addition to the calibration study for Enriched Curriculum researchers, the notes made by the researcher at the time of the Early Years Enriched Curriculum Evaluation Project observations were reviewed by Dr Walsh, in order to confirm that researchers were still acting in accord with the rubric which specifies the scoring within the QLI.

The same process was applied to the original data from the six early years experts, who validated the instrument and participated in the original calibration study. These data are presented in Table 4.

---

<sup>6</sup> Unlike some inter-rater co-efficients, it is not based on rank order. The larger the mismatch between any two raters, the greater the reduction in the co-efficient.

*Table 4*  
*Alpha (coefficient of inter-rater reliability) for all elements of the QLI:*  
*Data from six early years experts*

	Children's actions	Teaching strategies	Environment
Motivation	.83	.83	.69
Concentration	.73	.83	.69
Confidence	.73	1.0	.67
Independence	.83	.83	.83
Well-being	.83	.83	.83
HOTs*	.83	.73	.73
Multiple skills	.83	.83	.73
Social interaction	.73	.83	.73
Respect	.83	.83	.83

\* Higher order thinking skills

In all of the 27 scales, the data from the early years experts showed comparable or higher reliability than that of the student researchers. This is not surprising considering the much larger size of the Early Years Enriched Curriculum Evaluation Project researcher group: the greater the number of people asked to give a judgement the greater the probability of detecting variations in levels of concordance among them

A composite score was obtained by summing all 27 scores. There was no statistically significant differences between the scores of the early years experts and the student teachers in the calibration study.

The results of these calibration studies compare very favourably with similar studies on the ECERS (Sylva and Siraj-Blatchford 2001). The use of a single video for the calibration study for experts and non-experts alike and the use of a robust statistical method are superior aspects of the calibration of the QLI.

### **3.5 How costly are the available observational measures?**

In the case of this project, the evaluation team understood that certain quantitative measures relating to outcomes for children were an essential component. Further, it is expensive to gather these data because the work must be undertaken by trained testers with suitable background experience. Interview and questionnaire data could be gathered by the principal researcher on the project. The collection of further qualitative data is therefore circumscribed by the amount remaining in the budget.

Walsh's QLI, in addition to being a valid and reliable measure, incurred no expenditure on materials because it was developed by a member of the evaluation team. In addition, the researchers were willing to carry out the work as part of their professional development. Their relative inexperience meant that they were not ideal but it would not have been possible, within budget constraints, to undertake so many hours of classroom observation while paying more experienced personnel. Finally, Walsh's QLI had been

used extensively in a previous study, comprising observations in 38 classes suitable as controls for Enriched Curriculum classes. Considering the data available from the calibration study and all of the practical considerations, the evaluation team were satisfied that the observation study using the QLI would be a worthwhile exercise and would meet the criteria expected during peer review.

## 4. Cautionary notes

Our remit did not allow us to follow up the children in the second Shankill cohort, that is, those who would be experiencing the new curriculum under the aegis of teachers experienced in its implementation. Given the nature of the intervention and the knowledge that such a programme takes time to become fully established, the evaluation team would strongly recommend that the Enriched Curriculum is not judged solely on the data from the first cohort.

The evaluation does not address the question of long-term effects. If it should turn out that academic gains are negligible but learning dispositions are improved, the programme should be considered a success. Unfortunately, the effect of improvement only in learning dispositions is not likely to be reflected in later academic gains for some time. Such effects have been observed in many studies. For example, Head Start is a major internationally renowned programme which has demonstrated a loss of short-term academic gains in the medium term. Nevertheless, the long-term follow-up studies showed extensive gains in employment and social adjustment indicators in young adults who had experienced the intervention. Further information may be found on the many peer-reviewed studies reported at [www.nhsa.org](http://www.nhsa.org) and [www.acf.hhs.gov/programs/hsb/](http://www.acf.hhs.gov/programs/hsb/).

The final cautionary note relates to sample size. There is an inverse relationship between the effect size of an intervention and the size of the sample required to have a reasonable chance of detecting it. One cannot know the effect size in advance. However, reference to other educational interventions suggests that an effect size of one quarter of a standard deviation would be considered a resounding success. As the effect size falls below this figure, *although remaining statistically significant*, the chances of failing to detect it will rapidly increase. Although the effect size is still unknown, there is evidence to suggest that an improvement as large as one quarter of a standard deviation is somewhat unlikely. The evaluation team would recommend extending the sample size to test all the children in the cohort. At present, we do not have the funds to do so.

## 5. Conclusions

The selection of measures for inclusion in the assessment battery was made against four criteria:

- (i) Functionality: Will the instrument provide the type of information needed to answer the research question?
- (ii) Quality: Is the instrument valid and reliable?
- (ii) Developmental suitability: Is the instrument appropriate to the developmental levels of the school children and the practical requirements of the classroom?
- (iii) Cost effectiveness: Does the measurement instrument represent good value for money?

The first and second objectives of the evaluation study addressed the short and longer-term impact of the Enriched Curriculum on the learning dispositions and progress in literacy and numeracy of children against an appropriately matched control sample. The PIPS suite of age-appropriate tests was chosen as a cost-effective measure with high reliability and validity. PIPS measures achievements across a wide range of domains pertinent to the National Curriculum. PIPS is calibrated against the objectives of the National Curriculum and is not sensitive all of the intended learning outcomes of the Enriched Curriculum. It was augmented with a range of other attainment tests selected against the four criteria of functionality, quality, developmental suitability and cost effectiveness.

The third objective of the evaluation focused on recording the perceptions of teachers and parents regarding the appropriateness of the revised curriculum. Standard measures were not available. Bespoke interview schedules were devised. The quality of responses was strengthened by providing respondents with guarantees of confidentiality. The interview schedules were revised and updated in order to ensure that pertinent information was being recorded in a timely fashion.

The fourth objective of the evaluation focused on determining some of the factors that impact on the effectiveness of curriculum implementation, The evaluation rationale was to use a measure of classroom process in combination with a study in which a sample of children experiencing the Enriched Curriculum are systematically compared with a sample of controls. The QLI comprises an examination of children's learning experiences based on detailed classroom observation and analysis which has been successfully used to discriminate between Enriched Curriculum classrooms and a set of more traditional classrooms for which Northern Irish data are available. It was chosen because:

- It provides the type of information needed to answer the research question.
- It is valid and reliable.
- It is appropriate to the developmental levels of the school children and the practical requirements of the classroom.
- It represents very good value for money.

Approximately 85% of financial resources available for the evaluation are directed to the quantitative assessment of outcomes and 15% to observational assessments of the children's learning environment, parents' perceptions and teachers' perceptions. There is an inverse relationship between the effect size of the outcomes of the Enriched Curriculum and the size of the sample required to have a reasonable chance of detecting it. One cannot know the effect size in advance. Although the effect size is still unknown,

there is evidence to suggest that an improvement as large as one quarter of a standard deviation is somewhat unlikely. The evaluation team would recommend extending the sample size to test all the children in the cohort.

## 5. References

- Boehm Tests of Basic Concepts (2000). *The Boehm Pre-school Version & The Boehm 3*. London: The Psychological Corporation.
- Cowie, M. & Croxford, L. (1999). *Using Baseline Assessment for Early Intervention*. Centre for Educational Sociology, University of Edinburgh.
- Harms, T., Clifford, M. & Cryer D. (1998). *Early Childhood Environment Rating Scale: Revised Edition (ECERS-R)*. Teachers College Press
- Hogan, A.E., Scott, K.G. & Bauer, C.R. (1992). The Adaptive Social Behaviour Inventory (ASBI): A new assessment of social competence in high-risk three year olds. *Journal of Psychoeducational Assessments*, 10 (3), 230 – 239.
- Renfrew Language Scales (1969). *The Bus Story Test: A Test of Narrative Speech*. Bicester, Oxon: Winslow.
- Sylva, K., Siraj-Blatchford, I. & Taggart, B. (1999). (unpublished) *The Early Childhood Environment Rating Scales: 4 Curricular Subscales (ECERS-E)*. London: Institute of Education.
- Sylva, K. & Siraj-Blatchford, I. (2001). *The relationship between children's developmental progress in the pre-school period and two rating scales*. International ECERS Network Workshop, Santiago, Chile, July 2001.
- Tymms, P., Merrell, C. & Henderson, B. (1998). *Baseline Assessment and Progress During the First Three Years at School*. Paper presented at the conference of the British Psychological Society, Brighton, March 1998.
- Wechsler Objective Language Dimensions (1996). *WOLD Test of Written Expression*. London: The Psychological Corporation.

## **Appendix A**

### **The model for educational change in Northern Ireland**

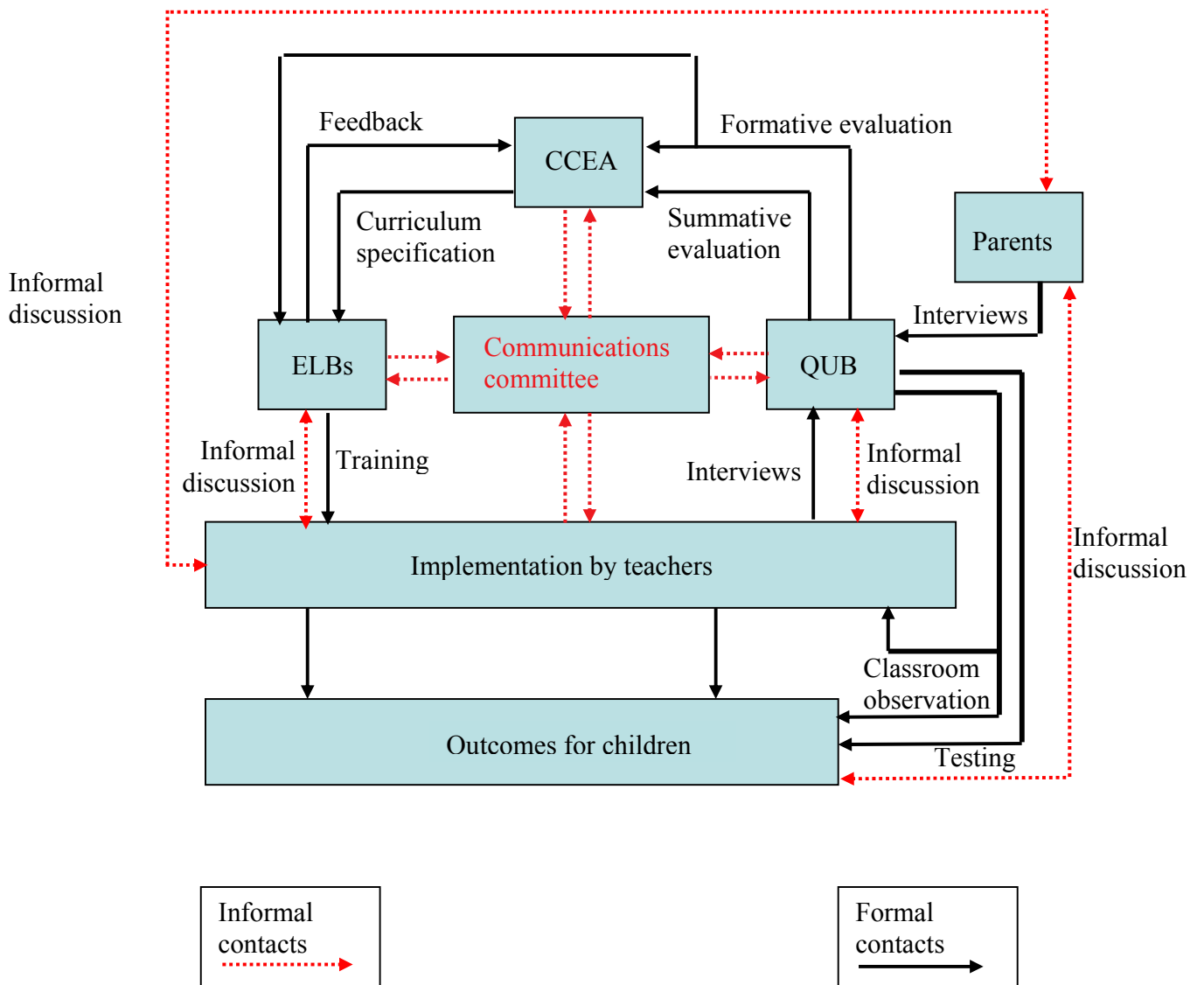
The origins of the Early Years Enriched Curriculum Evaluation Project were somewhat serendipitous. Following a meeting of staff and principals in 1999, educationalists involved with the Greater Shankill Area of Belfast came to the conclusion that the curriculum as it then stood did not sufficiently meet the needs of the children in the area. Staff in Belfast Education and Library Board (BELB) naturally became involved and were also working on the problem. Around the same time, early years personnel within the Northern Ireland Council for Curriculum, Examinations and Assessment (CCEA) were looking at research into early years education, especially into early years systems in other parts of the world. Thus, BELB, CCEA and the staff in schools found themselves to be working in the same direction. It was evident that it would be sensible to come together to formulate an alternative approach to early years education. This new approach came into being as the Enriched Curriculum. As the evaluation team made clear in their first report (Sproule et al. 2001), the new programme could be characterised at this time as an evolving curriculum. Indeed, this was seen as one of the strengths of the project because it provided an opportunity for all the professionals to have an input.

As the project was rolled out across the other Education and Library Boards (ELBs), the number of stakeholders increased rapidly. It now included all the relevant staff in the ELBs and the staff of the second set of schools. As the evaluation team had pointed out, in the absence of a written framework document for the Enriched Curriculum, there was no certainty that each ELB would interpret the Enriched Curriculum in exactly the same manner. A mechanism for promoting cohesion and for facilitating the evaluation across ELBs was evidently required. Accordingly, a communications committee, under the auspices of CCEA, was formed to fulfil these roles. The committee's first remit was to address any issue which might impede the implementation of the evaluation, such as negotiation of the testing timetable in order to minimise disruption in schools. It was soon addressing its second role by becoming a forum in which some of the findings of the evaluation could be shared informally among stakeholders, provided CCEA had given prior approval. By this means, there was a useful dissemination of those data that provided reassurance to teachers at a time of dramatic change for them. Crucially, it also provided a mechanism by which teachers, in particular, were able to make their voices heard, even to the extent of influencing the evaluation team and other professionals in appropriate ways. For the research team, the committee had many advantages. For example, it was a means of gaining the trust of teachers and obtaining their advice on such matters as the amount of work teachers might reasonably be expected to do as part of the evaluation. Gradually therefore, the communications committee became one of the agents for providing formative evaluation.

At regular intervals, the evaluation team were also providing formal reports to CCEA on the progress of the project. Some of this information has been used to inform future planning and to design information for parents and teachers.

In order to interpret the reports of the evaluation team to best advantage, it is helpful to consider how the relationships between the various stakeholders have come together to constitute a complex model for early years educational change in Northern Ireland. The model for change is presented in Figure 1 below. The various stakeholder groups in the model and their major roles are summarised as follows:

- CCEA
  - CCEA is a major source of funding for the project, including the funding (and commissioning) of the evaluation.
  - CCEA is responsible for agreeing a new, early years curriculum for Northern Ireland. This new curriculum will incorporate those aspects of the Enriched Curriculum for which there is solid evidence of efficacy. It will also reflect a consensus of opinion across the ELB early years experts.
- ELBs
  - ELBs organise the roll out of the project to new schools.
  - ELBs are also providing funding for the project. This funding mainly comprises money for teacher training within each ELB's own area.
  - ELBs are responsible for contributing their expertise to the preparation of the new early years curriculum. This takes place through working groups for experts in each of the curriculum domains.
  - ELB personnel are responsible for training teachers new to the project.
  - ELB personnel are responsible for assessing the efficacy of the programme in practice and providing feedback to CCEA and the evaluation team.
- Teachers
  - Teachers are responsible for delivery of the Enriched Curriculum.
  - Teachers are charged with providing the evaluation team with all the relevant information about their perception of the efficacy of the programme in practice.
  - Teachers are the most important group for informing parents.
- Parents
  - In today's society, parents are encouraged to take an active interest in their children's education and increasingly, to support schools in their delivery of that education. In the Enriched Curriculum project, parents are encouraged to take a particularly active role in supporting the work of the teacher at home.
  - Parents are encouraged to provide feedback to the evaluation team which in turn informs the further development of the project.
- The evaluation team
  - The evaluation team are an independent group from the School of Psychology, Queen's University Belfast and Stranmillis University College. This group is responsible for gathering quantitative and rich qualitative data from children, teachers, parents and other education professionals. These data are to be analysed in appropriate ways and presented to CCEA in a suitable format.
  - Initially, the evaluation team reports directly to CCEA. With CCEA's approval, formative feedback is released to other stakeholder groups, as and when that is deemed appropriate.



*Figure 1*  
The model for educational change

## **Appendix B**

### **Developing an Instrument to Assess a Quality Learning Experience in Early Years Settings**

**Dr Glenda Walsh**

Stranmillis University College

**Prof John Gardner**

Queens University Belfast

**ABSTRACT:** This paper details the development of an observation schedule which can be used to assess the quality of the learning experience in Year 1 classes. The instrument arose in the context of the debate about the relative importance of play-based and formal approaches to learning and teaching. The paper begins by presenting an outline of the features associated with an experiential model of young children's learning which forms the theoretical underpinning of the instrument. It then details the empirical development of the instrument, where the key features e.g. motivation, identified by the literature, were tested in context by conducting a series of observations in Danish kindergartens and Year 1 classes in Northern Ireland. Methods to ensure the validity and reliability of the instrument's usage conclude the article. Comments made by a team of early years specialists about the instrument are discussed and the results of the calibration study (undertaken to establish the inter-rater reliability of the instrument) are detailed.

### ***Introduction***

The instrument detailed in this paper, which we call the Quality Learning Instrument (QLI) arose from a research study which set out to examine the play versus formal debate relating to early years provision in Year 1 classes in Northern Ireland primary schools. Central to this study was the need to develop a means of assessing the quality of the learning experience. The Northern Ireland education system is perceived by some e.g. Caul (1990) to be too formal in approach, lacking the benefits thought to be offered by more play-based systems in other parts of the world. Kindergartens in Denmark were therefore used as an example of play-based approaches to early years education which could offer a comparative backdrop to the study.

Before an examination of the learning experiences could be undertaken, it was considered necessary to define what we mean by a high quality learning programme. Athey (1990) recognised this need, emphasising that descriptors of quality must precede any measurement process. She argued that "measurement without description and conceptual understanding can capture only the organisational surface of trivial features of situations" (p.8). Statham and Brophy (1992) also indicated that "any attempt to provide an objective rating scale for measuring quality has to assume that there is an explicit model of what

constitutes good provision” (p. 145). Instead of simply adopting an existing definition of quality against which to assess the programmes, Balageur, Mestres and Penn (1992) recommended that it is more appropriate to develop an in-house instrument, arguing that the “ the process involved in defining quality – with the opportunity it provides to explore and discuss values, objectives and priorities – is of utmost importance, and can be lost where people simply adopt existing measures” (p. 11). A major objective of the research was therefore to design, develop and test an instrument for assessing quality, which would then provide data for the central analysis of the study inasmuch as none of the off-the-shelf instruments were considered to be appropriate for the necessary data collection.

The intended focus of the instrument was on what Katz (1995) terms as a “bottom-up perspective of quality” i.e. how a programme is experienced by the participating children (p. 120). A review of existing assessment schedules suggested that the majority were test-bound i.e. the quality of the programme on offer is determined by measures of IQ and academic skills e.g. those used in studies such as that of Sundell (1992) and Tymms (1995). Others such as the Early Childhood Environment Rating Scale (Harms and Clifford, 1980 and Cryer 1996) tend to focus more on a “top-down perspective of quality” (Katz, 1995, p. 120) such as the accessibility of the provision, the curriculum, the equipment, the child-staff ratio; and not on the actual quality of the learning experience itself. The alternative chosen in this study was based on identifying the framework for an instrument from the literature. This basic framework was then developed and trialled to ensure its validity in the settings under investigation.

### ***Theoretical Underpinning of the Instrument***

The QLI draws on a model of how young children learn, a model which might be summarised as experiential. This model of learning draws heavily on the earlier work of philosophers such as Dewey (1938) who expressed the belief that “all genuine education comes through experience” (p. 25). Piagetian ideas are also acknowledged in this model of learning, where children are believed to construct their own knowledge through interaction with the environment. The child is not however perceived to learn in isolation but rather in the company of significant others who can support them as they learn. In this way the experiential model of learning is also deeply rooted in the Vygotskian notion of social constructivism.

Six key features of the experiential model underpin the QLI, namely that:

- children should be actively engaged in their learning;
- children need some control and autonomy over their own learning;
- children must feel secure in their learning environment;
- children learn in the company of others;
- children’s learning is holistic rather than based on subject areas; and
- children’s metacognitive powers must also be considered.

### **Active Engagement**

Young children need to engage actively in the learning process to ensure effective

learning takes place. As Watson (2000) states “knowledge is not passively received and absorbed but actively built up by the individual” (p. 136). Laevers (1993) referred to this mental activity as a source of intense involvement which facilitates the overall development of the young child. He defined involvement as “a quality of human activity, characterised not only by a high level of motivation, but also by concentration and persistence, intense perceptions and experience of meaning, a strong flow of energy and a high degree of satisfaction” (p. 61).

Proponents of the experiential model therefore recognise the importance of intrinsic motivation for young children’s learning and educational achievement. Psychologists (e.g. Dweck and Leggett, 1988, Gottfried, 1990 and Ames, 1992) stress the need to initiate in children a desire to want to learn. They become what Dweck (1986) referred to as “mastery” learners - i.e. learners who are challenge-seeking, who persist in the face of difficulty and who enjoy “exerting effort in the pursuit of task mastery” (p. 1040).

Inherent, therefore, within the experiential model is the idea that fostering a positive disposition towards learning (i.e. developing an environment in which children are fully motivated and actively absorbed in the learning process) is as important as developing young children’s knowledge and skill acquisition (Katz, 1995, 1999).

#### *Autonomy*

Embedded in much of the experiential literature is the belief that children should have some control over the learning that takes place. Howe (1984) took this point of view, suggesting that children’s independent actions and feelings of self-control are important to later development. He argued that when children believe that the outcome of a situation depends on their own actions, they engage more effort in the process.

#### Emotional Security

A wealth of convincing evidence (e.g. Greenhalgh, 1994; Goleman, 1996 and Laevers, 1996, 1999) has referred to the importance of children’s emotional stability for learning and development. For example Goleman (1996) reported that people in general possess two different kinds of intelligences, namely rational and emotional, which affect their potential in life. He indicated that people with well-developed emotional skills are more likely to be content and effective in their lives.

#### Social Interaction

The need for positive social relationships has been identified as another feature of an experiential learning environment. Vygotsky (1926 in translation, 1978) advocated the importance of rich interactive settings for profitable learning experiences. He emphasised the crucial role of significant others in children’s learning, in helping the children to extend their learning beyond what they can do alone. Rogoff (1990) supported Vygotsky’s thinking, emphasising:

*“...day to day engagement of children and adults in shared activities contributes to the rapid progress of children in becoming skilled participants in the intellectual and social lives of their society.... like genes, social interaction and social arrangements are an essential aspect of child development, without which it would be impossible to conceive of a child developing” (p. 138).*

### Children Learn Holistically

Another integral feature of the experiential model is that children’s learning is not separated into distinct subject areas, but rather that it is holistic in nature. This view is expressed by Gardner (1993, 1999) who advocated the importance of a broad and balanced curriculum. He argued that children possess ‘multiple intelligences’, comprising linguistic, logical mathematical, musical, spatial bodily kinaesthetic, interpersonal and intrapersonal skills, all of which, in his view, require enhancement. Experiential proponents are of the opinion that young children are capable of demonstrating sophisticated powers of complex thinking when provided with an appropriate learning environment (see e.g. Aubrey’s research, 1993 on the mathematical competency of young children), and for this reason advocate what Katz (1995) referred to as “educative” (p. 90) experiences, rather than “frivolous one shot activities” (p. 35).

### Metacognition

Based on an experiential model, a high quality learning environment needs to develop children’s metacognitive powers and skills. According to Meadows (1993), these skills include “analysing and defining the character of the problem at hand; reflecting upon one’s own knowledge (and lack of knowledge) that may be required to solve the problem; devising a plan for attacking the problem; checking how the plan helps in the problem-solving.... and, generally, orchestrating cognitive processes in relation to the cognitive contents and the objectives involved” (p. 78/79). In this way it is important that the learning environment enables young children to think about what they are doing and encourages them to engage in a reflective process, which develops their higher order thinking skills.

A high quality learning environment from an experiential perspective can be therefore summed up in the words of Adams (1996):

*“A quality curriculum addresses not just the moment, but prepares children for the next step ahead. The best way to prepare children for tomorrow is to give them a good today - a today that is challenging but not threatening or intimidating. A today that encourages children to think of themselves as learners and to accept and appreciate those around them” (p. 52).*

Drawing from this experiential model of learning, seven key features were therefore considered integral to the development of the Quality Learning Instrument (QLI) namely:

- motivation
- concentration
- confidence
- independence
- social interaction
- multiple skill acquisition
- higher order thinking skills

## ***The Pilot Observation Studies – The Empirical Evidence***

Although the development of the QLI was embedded in an experiential model of learning, its formulation was also mediated by evidence from a series of pilot observation studies, the views of early years experts and a calibration study.

The pilot observation studies were carried out in eight Danish kindergartens and in ten Year 1 classes in Northern Ireland. The observations were conducted over a period of two full days i.e. 9am –2pm in the Northern Ireland Year 1 classes and 8am –6pm in the Danish kindergartens.

The aim of the pilot observations was to evaluate the way in which the key features of quality learning were in operation in actual early years settings and to identify examples that indicate high or low quality experiences. The features themselves were not intended to be measured, but evaluation of their manifestation was designed to provide an indication of the overall quality of the learning experience in different early years settings.

During each of the observation visits, the intention was to gather as much data as possible against a triangle of interaction: children, teacher and the environment, based on the thinking that children's learning does not take place in isolation (see e.g. Vygotsky, 1978; Brofenbrenner, 1989 and Dalhberg, Moss and Pence, 1999). The format of the observations involved note-taking throughout the visit, structured in terms of processes and interaction around the children and teachers and including descriptive details of the environment, set against the key features identified from the literature survey. The observations produced large volumes of field notes taken in situ, which required analysis before leading to instrument design. A form of thematic analysis was used. In this method pre-selected themes were used to condense the data. The data was categorised in a matrix format by selecting observations relating to the themes e.g. motivation, independence, happiness etc. and the vertical axis listed the triangle of interaction i.e. children's actions, teaching strategies and the role of the environment. Each matrix cell therefore, contained evidence which grouped the theme with the interaction focus e.g. the children's actions in the context of their motivation.

The next stage in the data reduction process used a further matrix in which the themes were listed against the observation settings. In this way, the matrix held all of the selected data from the series of observations and each matrix cell contained evidence against a particular theme in accordance with a particular setting. The following table (A) illustrates the process with examples from different research sites in the form of details of the hand-written field notes from the pilot stage of observations:

Table A: Illustration of thematic analysis matrix for the pilot observations

Research Sites	Motivation	Concentration	Independence	Themes continued
1	Activities on offer are stimulating and practical. Children are very keen at story and playtime. Environment is cheerful and colourful	Few of the children appear distracted. Children are mainly involved in what they do, showing some precision in the process.  On occasions the teacher appears to be challenging the children.	On a few occasions the children are encouraged to participate in classroom chores. Few signs of initiative being shown. Activities appear quite directed. Teacher decides what should be done and when. The furniture is child-sized but children are not free to use the materials unless the teacher tells them to do so.	→
2	Children appear bored. The majority have dull expressions. Little eagerness is shown. Activities appear quite boring and repetitive. There are few opportunities for hands-on work. The environment is dull.	Teacher is in control and maintains a level of concentration by walking around the classroom and reprimanding anyone who does not work. On occasions the children are engaged in time wasting activities but very quickly in most instances they are brought back on task by the teacher.	Children have no choice. They are told exactly what to do. Specific time set aside to go to toilet. Some children are encouraged to deliver messages and to help at tidying up	→
Research sites continued	↓	↓	↓	↓

The next stage of analysis involved consulting the data matrix to categorise the data as ‘high’ to ‘low’ examples of each theme. An example from the results which came from this process is provided in the following tables for the key theme motivation.

### Table B: Examples of ‘high’ levels of motivation

#### Motivation

The observations suggested that a high level of motivation was in evidence when:

- the children were eager to participate in the activities;
- they were energetic, enthusiastic and they displayed a degree of curiosity and interest in the activities
- adults offered stimulating, relevant and age-appropriate activities;
- they showed a degree of interest and interacted appropriately, allowing the children some degree of freedom and choice;
- they were cheerful and enthusiastic.
- the environment was spacious, airy and aesthetically pleasing;

#### Examples of high motivation:

- 1) In kindergarten (5) (Denmark, hereafter known as D) many children spent most of the afternoon outdoors. Some were running through the bushes, pretending to be cowboys and Indians. Others were busy helping a pædagog in the greenhouse, or watching in anticipation as a pædagog lit a campfire. Another group was splashing about in the sandpit, which was saturated with water, getting themselves as wet and dirty as possible.

- 
- resources tended to be plentiful, attractive and age-appropriate;
  - some exciting areas were available e.g. an Aladdin's cave reading corner, a cellar;
  - children got the opportunity to use their environment, both inside and outside.
- 2) In classroom 4 (Northern Ireland, hereafter known as NI), during structured play, a group of children were eagerly looking at photographs of their trip to the zoo. After chatting about their experiences, they painted pictures of the animals they saw to add to the zoo display.
  - 3) A group of five boys were playing on trucks in kindergarten 2 (D). They were making the sound of a fire siren and added a rope and some buckets to their trucks. A pædagog provided them with a hose and they pedalled hastily to the sand tray to put out 'the bush fire'.
- 

*Table C: Examples of 'low' levels of motivation*

The observation data suggested that the level of motivation was low when:

- 
- |  |  |
|--|--|
| <ul style="list-style-type: none"> <li><u>-the children appeared apathetic and unenthusiastic e.g. lying over the tables, wandering around the room, yawning etc.;</u></li> <li><u>-they seemed to complete the activity out of obligation rather than interest.</u></li> <li><u>-adults either showed little interest in the children's activities or totally dominated them;</u></li> <li><u>-activities were uninteresting, neither age-appropriate nor relevant to young children;</u></li> <li><u>-little variety was offered and little or no choice was given;</u></li> <li><u>-the environment was dull and lacking in character;</u></li> <li><u>-resources tended to be routine and uninspiring;</u></li> <li><u>-space was limited;</u></li> <li><u>-children had little opportunity to use the environment available.</u></li> </ul> | <p><b><u>Examples of low motivation:</u></b></p> <ol style="list-style-type: none"> <li>1) <u>The children in classroom 8 (NI) had been asked to complete a worksheet, which involved them colouring a snake red, to allow the teacher to hear reading. Having completed the activity quickly, easily and in many cases carelessly, a group of boys began to hide under tables and throw books across the table at one another.</u></li> <li>2) <u>In kindergarten 2 (D), the children, before lunch, were standing in a circle, singing songs, which did not seem to appeal to the children. Several children started to fidget and did not participate fully.</u></li> <li>3) <u>In classroom 1 (NI) the entire class was involved in playing the 'farmer wants a wife' in the assembly hall during a PE lesson. The children sang the song repeatedly (approximately five times) showing little signs of</u></li> </ol> |
|--|--|
-

---

enthusiasm in the process. A group of boys started to pull the others in the circle and then ran to the toilets. Other children then left the circle and went to the toilet.

---

The pilot observation studies also suggested that some reference to the children's physical well-being, defined as contentment, comfort (including healthy eating and safety), is requisite in any description of a high quality learning environment. Similarly the observations suggested that respect for others (i.e. child to child, adult to child and child to adult) was a feature of any high quality learning environment. High and low definitions for each of these new key themes were also included.

The next refinement in the instrument design involved presentation. To allow for easy reading, each key indicator was placed on a separate page, with high and low descriptors for each key theme. These related to the children's actions, teaching strategies and the role of the environment allowing the researcher to view the already gathered evidence with relative ease before making his/her overall decision about the quality of the setting.

The final stage of the instrument design was to validate it and the next section outlines this process.

## **Assuring the Validity and Reliability of the QLI**

### The Views of Early Years Experts

The next stage in the development process was to send the QLI to a group of early years experts ( eight in Northern Ireland and six in Denmark) to comment on its validity. The sample of experts included: one government inspector, five university lecturers, two local authority advisors, two early years researchers and four early years teachers in management positions. The average number of years experience was approximately 23.

The sample of experts was all in agreement that the QLI addressed key indicators of quality practice in the early years. Overall they expressed the view that the examples relating to each indicator were relevant and comprehensive. The Danish experts expressed their satisfaction with the way in which the schedule referred to skill areas other than reading, writing and arithmetic. Both the Danish and Northern Ireland experts agreed that the format was simple and straightforward.

The final collection of features, categorised along the axes of child, adult and environment, was then ready for use as the Quality Learning Instrument, QLI.

As the criteria and the judgements made when using the QLI were based on only one judge some form of calibration had to be undertaken.

### ***Calibration of the Instrument***

As a single judge of the quality of the learning experiences on offer, the research judgements are open to criticism on the grounds that an objective reflection cannot be guaranteed. Despite the researchers' expertise in the field of early childhood education and the video and paper-based accounts to support the judgements made, the question is still begged how reliable are the judgements? (i.e. would other judges record the same findings?). Kirk and Miller (1986) used the analogy of a thermometer to explain this process, describing reliability as "the extent to which a measurement procedure yields the same answer however and whenever it is carried out" (p. 19).

For this reason a calibration study was conducted to ensure the judgements were as objective as possible. The ten raters who participated in the calibration process were all practising early years teachers, the majority of whom held a position of responsibility i.e. two were vice-principals and four were Key Stage 1 co-ordinators. No male teachers participated in the study.

Each of the raters was sent an extract of video, taken in a Danish kindergarten, accompanied by a set of instructions, some photographs and a copy of the QLI. These requested that the video should be viewed initially, followed by consultation of the QLI. The video was then to be re-watched and in the grid provided, examples of practice were to be noted. The teachers were asked to rate the children's actions, the teaching strategies and the environment on offer using the QLI as the observation instrument. Each category i.e. children's actions etc. in each of the targeted areas, was then to be scored on a 5-point scale i.e. 5 being at the highest end and 1 at the lowest of a spectrum of high and low quality learning activities. The video was then to be viewed for a third time and any further comment was to be added. A sample record sheet based on observation in an imaginary kindergarten was included in the pack to provide the teachers with some concrete evidence as to what was expected.

The video excerpt was also viewed and assessed by the researchers in advance but the raters were not aware of their scores. The calibration process involved comparing the researchers' scores to those of the teachers to establish whether the researchers' scores were of the same size and direction.

### *The Teachers' Scores*

Although some teachers might only give full marks (i.e. a score of '5') in exceptional circumstances, while others might be of the opinion that a '5' is reflective of good quality practice, the raters' and the research scores were very similar. In relation to each category (i.e. children's actions etc.) of motivation, concentration and confidence, there was never less than seven of the teacher raters in agreement. For those who were in conflict with the others it was always a single score difference e.g. '5' instead of '4'.

The research scores also corresponded highly with the raters' mean scores for each of these features as is illustrated in table D. The scores were multiplied by 20 to give a score out of 100 for ease of reading.

*Table D: Raters' mean scores and the research scores for motivation, concentration and confidence:*

Feature	Motivation			Concentration			Confidence		
	CA*	TS	E	CA	TS	E	CA	TS	E
<b>Mean score</b>	<b>84</b>	<b>76</b>	<b>64</b>	<b>86</b>	<b>76</b>	<b>66</b>	<b>86</b>	<b>64</b>	<b>82</b>
<b>Researcher</b>	<b>80</b>	<b>80</b>	<b>60</b>	<b>80</b>	<b>80</b>	<b>60</b>	<b>80</b>	<b>60</b>	<b>80</b>

\*CA = Children's Actions

TS = Teaching Strategies

E = The Role of the Environment

Similar levels of agreement between the teacher raters were achieved for all the other indicators such as independence, social interaction etc. with the exception of the environment. Due to the high level of agreement in the other areas of practice, it was considered more likely that the disparity of opinion on the environment issue might be explained by the lack of adequate evidence shown on the video extract. Perhaps some of the raters concentrated more than others on the additional background information and photographs, which might have influenced their findings. Furthermore the environment is seldom taken into consideration when evaluating the quality of the learning experience and for this reason some of the teacher raters may not have been familiar with this practice.

Nevertheless in all cases the extent of overall agreement between the raters' mean score and that of the researchers was very strong.

The results from the calibration study allowed the researchers to consider that their judgements arising from the use of the QLI were reliable in early years settings as the high level of agreement obtained on most of the features indicated a strong inter-rater reliability. The security of this deduction is strengthened by the fact that the raters were unknown to each other and came from different schools. They also had different levels of Year 1 experience and in some cases training. Furthermore they completed the task individually by post and had no indication of the researcher's or each others' ratings.

## Conclusion

Although not original in the sense that it was derived from other sources, the Quality Learning Instrument (QLI) possesses a degree of originality in the way in which it crossed two cultures and was calibrated by a number of early years experts. It could be argued that this instrument has the potential to be used as an assessment schedule in any early years setting, taking into consideration the cultural diversities that exist.

The content of the QLI, based on nine indicators of quality learning e.g. motivation, concentration etc. bear some resemblance to the four outcome areas that Pascal and Bertram (1999) have identified for their AcE Project (Accounting Early for Life Long Learning). In this way it could be argued that the indicators used in the QLI not only act as process indicators but could also be used as outcome measures as Pascal and Bertram (1999) argue that it is time to stop thinking of outcomes simply in terms of “facts, subjects and disciplines of knowledge” (p. 101/102). Laevers (2000) also argues this to be the case, challenging the view that “narrowly defined academic achievements are the only means of measuring educational outcomes” (p. 20). Perhaps the QLI could supplement the work of Pascal, Bertram and Leavers in this field as well as frameworks such as the “Quality in Diversity” (Early Childhood Education Forum, 1998) or Carr’s “dispositional framework” (Carr, 1998) all of which emphasised the importance of positive dispositions as measures of learning outcomes. In fact it could be argued that the QLI provides not only an easy to use and comprehensive assessment schedule for research purposes but also a means for early years teachers to assess the quality of their own practice and inform and develop their understanding of children’s learning.

The QLI is now being used as one of the main assessment instruments in the Early Years Enriched Curriculum Evaluation Project (Sproule, L., Trew, K., Rafferty, H., Walsh, G., McGuinness, C. and Sheehy, N., 2001), a longitudinal study which aims to evaluate the quality of an innovative play-based curriculum, which is being trialled in a number of Year 1 classes in Northern Ireland. In due course the instrument itself will be further evaluated and refined in the light of these extensive trials.

## References

- ADAMS, L. (1996) Quality curriculum for quality care and education, *OMEP: International Journal of Early Childhood*, 28 (2) pp. 49-53.
- AMES, C. (1992) Classrooms: goals, structures, and student motivation, *Journal of Educational Psychology*, 84 (3) pp. 261-271.
- ATHEY, C. (1990) *Extending Thought in Young Children. Parent-Teacher Partnership* (London, Paul Chapman).
- AUBREY, C. (1993) An investigation of the mathematical knowledge and competencies which young children bring into school, *The British Educational Research Journal*, 19 (1) pp. 27-41.
- BALAGEUR, I., MESTRES J. AND PENN, H. (1992) Quality services for young children, *ERIC Document* (Washington, EDRS, ED 346957).
- BERTRAM, T. AND PASCAL, C. (1995) Questions of quality, in: Gammage, P. and Meighan, J. (Eds) *Early Childhood Education: The Way Forward* (Derby, Education Now).
- BROFENBRENNER, U. (1989) Six theories of child development, in: Vasta, R. (Ed) *Annals of Child Development*, Vol 6 (London, JAI Press, Inc).
- CARR, M. (1998) *Learning Stories: A Framework for Assessing Children’s Experiences in an Early Childhood Setting*. Paper delivered at the 8<sup>th</sup> European Conference on Quality in Early Childhood Research in Santiago de Compostela, Spain, September.

- DAHLBERG, G., MOSS, P. AND PENCE, A. (1999) *Beyond Quality in Early Childhood Education and Care* (London, Falmer Press).
- DECI, E. L. & RYAN, R. M. (1980) The empirical exploration of intrinsic motivational processes, in: Berkowitz, L. (Ed) *Advances in Experimental Social Psychology*, Vol 13 (New York, Academic Press).
- DECI, E. L. AND RYAN, R., M. (1985) *Intrinsic Motivation and Self-Determination in Human Behaviour* (New York, Plenum Press).
- DWECK, C. S. (1986) Motivational process affecting learning, *American Psychologist*, 41 (10) pp. 1040-1048.
- DWECK, C.S. & LEGGETT, E.L. (1988) A social-cognitive approach to motivation and personality, *Psychological Review*, 95 pp. 256-273.
- EARLY CHILDHOOD EDUCATION FORUM & THE NATIONAL CHILDREN'S BUREAU (1998) *Quality in Diversity in Early Learning. A Framework for Early Childhood Practitioners* (London, National Children's Bureau Enterprises).
- GARDNER, H. (1993) *Multiple Intelligences: The Theory and Practice* (New York, Basic Books).
- GARDNER, H. (1999) Intelligence reframed: multiple intelligences for the 21<sup>st</sup> century, *ERIC Document* (Washington, EDRS, ED 435 610).
- GOLEMAN, D. (1996) *Emotional Intelligence. Why It Can Matter More than IQ* (London, Bloomsbury).
- GOTTFRIED, K. E. (1990) Academic intrinsic motivation in young elementary school children, *Journal of Educational Psychology*, 82 (3) pp. 525-538.
- GREENHALGH, P. (1994) *Emotional Growth and Learning* (London, Routledge).
- HARMS, T. & CLIFFORD, R. (1980) *The Early Childhood Environment Rating Scale* (New York, Teachers College Press).
- HARMS, T., CLIFFORD, R. & CRYER, D. (1996) *The Early Childhood Environment Rating Scale Revised* (London: Teachers College Press).
- HOWE, M. J. (1999) *A Teacher's Guide to the Psychology of Learning* (Oxford, Blackwell).
- KATZ, L. (1995) *Talks with Teachers of Young Children* (New Jersey: Ablex Publishing Co-operation).
- KATZ, L. (1999) Another look at what young children should be learning, *ERIC Document* (Washington, EDRS, ED 430 735).
- KIRK, J. & MILLER, M. L. (1986) *Reliability and Validity in Qualitative Research* (London, Sage Publications).
- LAEVERS, F. (1993) Deep level learning: an exemplary application on the area of physical knowledge, *European Early Childhood Education Research Journal*, 1 (1) pp. 53-68.
- LAEVERS, F. (1996) *Social Competence, Self-organisation and Exploratory Drive and Creativity: Definition and Assessment*. Paper delivered at the 6<sup>th</sup> European Early Childhood Education Research Association Conference on the Quality of Childhood Education in Lisbon, September.
- LAEVERS, F. (2000) Forward to basics! Deep-level learning and the experiential approach, *Early Years*, 20 (2) pp. 20-29.
- MEADOWS, S. (1993) *The Child as Thinker* (London, Routledge).
- PASCAL, C. & BERTRAM, T. (1999) Accounting early for lifelong learning, in: Abbott, L. & Moylett, H. (Eds) *Early Education Transformed* (London, Falmer Press).
- ROGOFF, B. (1990) *Apprenticeship in Thinking* (Oxford, Oxford University Press).
- SPROULE, L., RAFFERTY, H., TREW, K., WALSH, G., MCGUINNESS, C. & SHEEHY, N. (2001) *The Early Years Enriched Curriculum Evaluation Project: First Year Report* (Belfast, School of Psychology, Queen's University Belfast).
- STATHAM, J. & BROPHY, J. (1992) Using the early childhood environment rating scale in playgroups, *Educational Research*, 34 (2) pp. 141-148.
- SUNDELL, K. (1992) Instructional style, age span in childhood groups and speech, cognitive and socio-emotional status, in: Laevers, F. (Ed) *Defining and Assessing Quality in Early Childhood Education*, Studia Paedagogica (Leuven, Leuven University Press).
- TYMMS, P., MERRELL, C. & HENDERSON, B. (1995) The reception year: a quantitative investigation of the attainment and progress of pupils, *Unpublished Paper* (Durham, Durham University).
- VYGOTSKY, L. S. (1978) *Mind in Society: The Development of Higher Mental Processes* (Cambridge, Harvard University Press).
- WATSON, J. (2000) Constructive instruction and learning difficulties, *Support for Learning*, 15 (3) pp. 134-141.