

Observation Methodology in Preschool Programme Evaluation: The Early Start Experience

Mary Lewis and Peter Archer

The ECERS-R, a classroom observation otherwise known as the Early Childhood Environmental Rating Scale, was used as part of the evaluation of Early Start, a preschool programme in disadvantaged areas in Irish cities. An evaluation report to the Department of Education and Science (Lewis and Archer, 2003) is concerned with what was learned about Early Start from the use of the ECERS-R. In addition to summarising that report, the present paper focuses on what can be learned about the ECERS-R from the experience of its use in the Early Start evaluation.

Early Start and its Evaluation

As part of an integrated approach to problems of disadvantage in designated schools, Early Start was established in eight locations in the 1994/95 school year and in a further thirty-two locations in the following year, to provide for three-year old preschool children. Though broadly concerned with the development of the whole child, Early Start was primarily designed to promote language and cognitive development. It has a number of features that differentiate it from provision for the infant classes of primary schools, including a shorter school day and a class size that is limited to fifteen pupils. Also, each Early Start teacher is assisted by a full-time childcare worker and has the support of a Home-School-Community Liaison (HSCL) Co-ordinator. The role of the HSCL Co-ordinator is to promote parent involvement and integration of the school with other local agencies and individuals involved in dealing with disadvantage.

An evaluation of Early Start in the original group of eight participating schools was undertaken during its first four years of operation. Data on implementation obtained in questionnaires and achievement data were examined (Educational Research Centre, 1998; Kelly and Kellaghan, 1999). Baseline achievement data were obtained in tests administered to pupils who were attending junior infants and second class when Early Start was introduced. The test results were subsequently compared with those of the first two cohorts of pupils that included Early Start graduates when they reached junior infants and second class. Interviews with junior-infant teachers provided additional information on pupils' adaptation to school.

The evaluation yielded mixed results. Early Start participants, when they reached junior infants, were judged by their teachers to have higher levels of cognitive and social maturity, to be better adapted to classroom procedures, and to have higher levels of self-determination and independence than their peers who had not participated in Early Start.

In literacy and numeracy, however, the achievements of the first two cohorts of Early Start pupils in junior infants were not found to differ significantly from those of pupils who had not attended Early Start, although language performance of the second cohort was significantly better than that of the first cohort (Educational Research Centre, 1998). Similar results emerged from the tests of reading and mathematics involving second class pupils (Kelly and Kellaghan, 1999). Differences in the achievements of pupils who had attended Early Start and those who had not, were not found to be significant.

Though not inconsistent with the findings of evaluations of early childhood interventions in other countries, the results of the Early Start evaluation were regarded as disappointing. In a review of the findings (Kelly and Kellaghan, 1999), it was suggested that implementation problems may have contributed to the failure of Early Start to impact on achievement. A number of these were targeted in a series of measures introduced by the Department of Education and Science (DES) in the late 1990s. The measures included additional in-service support involving visits to classrooms, the preparation of a curriculum with specific objectives, an increased emphasis on adult/child interactions, and the development of assessment profiles. Further evaluation was conducted to investigate the extent to which change had occurred in these and other aspects of implementation that had been identified as problematic.

The report that followed concluded that a good deal of change had occurred in some, but not all, aspects of Early Start since completion of the previous evaluation (Lewis and Archer, 2002). However, since the report's findings were limited to questionnaire data and an analysis of documents (curricular guidelines, teacher notes), it was agreed that some further investigation of implementation would be required and it was decided to proceed with an observation study. The decision was supported by the considerable progress that had occurred in the development of observation instruments for evaluation of early childhood settings (Essa and Burnham, 2001). Several such instruments were published in the latter half of the 1980s and the 1990s (e.g., Bredekamp, 1986; Howes and Stewart, 1987; Arnett, 1989; Abbott-Shim and Sibley, 1992; File and Kontos, 1993; McCartney *et al.*, 1997). A search was undertaken to establish if one of these might be applied or adapted for use in the Early Start evaluation, thereby creating an opportunity to link developments in Irish preschool education with those in other countries.

The Instrument

A revised version of the ECERS (Harms and Clifford, 1980), known as the ECERS-R, was identified in the review. According to Harms *et al.* (1998:1), who carried out the revisions, the ECERS-R is based on a "*...broad definition of environment, including those spatial, programmatic, and interpersonal features that directly affect the children and adults in an early childhood setting.*" These features are reflected in the scale, which

consists of seven subscales: space and furnishings, personal care routines, language-reasoning, activities, interaction, programme structure and parents and staff.

In the literature on childcare quality, the environmental rating scale approach is viewed as part of an increasingly sophisticated body of work that is designed to measure process quality as distinct from structural quality. While the latter is concerned with those elements of care which focus on describing the framework of a programme (e.g., teacher-child ratio, teacher qualifications), process elements of quality focus on the quality of interactions between caregiver and child and the quality of activities available for children. In their comprehensive review of the literature, Essa and Burnham (2001:73) conclude that it is the process variables which "*...seem to encompass the most important elements of a high quality program.*" The case for using an instrument such as the ECERS-R to examine the process quality aspects of Early Start was strengthened by the fact that much was already known about the structural aspects of the programme.

A number of specific characteristics of the ECERS-R also pointed towards the appropriateness of the scale for Early Start observation. First, the instrument has a strong educational focus and seemed conceptually congruent with the educational ethos of Early Start. Second, each scale item is supported by a set of indicators to assist scoring, as well as a series of questions, which seemed particularly useful in the case of non-observable behaviours, that might be put to teachers and childcare workers. Third, the instrument is designed to facilitate observation of classroom units rather than of individuals within classrooms. A fourth important characteristic of the ECERS-R is that it has been used in a variety of settings over a number of years. While most of this work was conducted in settings in the United States, the instrument has also been used and extended in a major longitudinal study of 141 preschool centres involving over 3,000 children in the UK (Sylva *et al.*, 2003). Finally, the ECERS-R can be administered within a period of hours, and was considered feasible from a resource point of view, since it seemed likely that it could be administered during one session.

Several modifications to the scale were applied before it was administered in Early Start centres. First, a number of items were deleted because of an apparent lack of fit with Early Start objectives. For example, items on play with sand, water, and blocks were excluded because the emphasis in the ECERS-R is largely on the availability of these resources rather than on their use as aids to learning as prescribed in the Early Start guidelines (In-Career Development Team, 1998). Next, the guidelines were inspected with a view to identifying omissions in the scale. An item on oral language development was added to reflect the complexity of the dialogue strategies promoted in Early Start and one on cognitive development was largely rewritten, also with reference to Early Start objectives. Finally, other items were amended (e.g., some indicators and questions were

altered, added or deleted) before the modified scale was submitted for review to a member of the Early Start Department of Education and Science team and to an experienced independent observer of Early Start centres.

In the scale that was used in Early Start centres, all seven of the ECERS-R subscales were retained but the number of scale items was reduced from forty-three to twenty-eight. Items were allocated to the subscales as follows: space and furnishings (eight items); personal care (three items); language-reasoning (three items); activities (four items); interaction (four items); programme structure (three items); and parents and staff (three items).

Fieldwork

Twenty of the forty schools involved in Early Start were randomly selected for observation visits in May and June of 2002. School principals were informed of the proposed visit and consulted on the date of the visit. The request was declined in one case for reasons relating to staff turnover.

Observations were conducted independently by two individuals (one visited twelve schools and the other visited seven) who collaborated to ensure consistency in the interpretation of item indicators and in the application of scores. In addition to completing the observation schedule, the observers interviewed the Early Start teacher in all cases and in the majority of cases, depending on the nature of classroom activity and the time available, the childcare worker. Interaction was limited between the children and the observers but, otherwise, the observation was conducted in accordance with the recommendations issued by Harms *et al.* (1998:5). The average length of visits was approximately two-and-a-half hours.

Following the fieldwork, the items were scored using the score sheets developed for the ECERS-R. Each item was rated on a scale of one to seven (inadequate =1, minimal =3, good =5, and excellent =7) depending on the number of indicators ticked during observation. This meant that the maximum score that could be assigned to a centre is 196 (28 x 7). In some centres, however, no score could be assigned to an item on disability since no pupil with a disability was enrolled, and in these cases the maximum score is 189 (27 x 7).

Results

Scale and Subscale Scores

Looking first at the overall score and the subscale scores, it may be noted that the maximum scale score that could be obtained across all centres is 3,724. However, the total number of score points allocated was 2,405, or just 64.58% of the maximum scale

score (Table 1). At subscale level, the highest score was achieved for items on interaction to which 84.58% of the maximum subscale score was allocated. The next highest scores related to language-reasoning (71.42%) and space and furnishings (68.42%), followed by activities (60.15%) and personal care routines (59.89%). The lowest scores were associated with the parents and staff subscale (49.87%) and the programme structure subscale (46.11%).

Table 1: Maximum and Achieved Scores in Early Start Centres (N=19), by Subscale

Subscale	Maximum Score	Achieved Score	Achieved Score as % of Maximum Score
Space and furnishings	1,064	728	68.42
Personal care routines	399	239	59.89
Language-reasoning	399	285	71.42
Activities	532	320	60.15
Interaction	532	450	84.58
Programme structure	399	184	46.11
Parents and staff	399	199	49.87
Total	3,724	2,405	64.58

Individual Item Scores

Though space constraints do not permit presentation of the data (see Lewis and Archer, 2003), an examination of scores at item level indicates variation in the distribution of item scores within subscales. In two of the subscales, scores are evenly distributed across items (relatively high scores on all items in the interaction subscale but only average scores on all items in the activities subscale). A bipolar distribution pattern, high scores on some items but not on others, is marked in the subscales on space and furnishings, personal care routines, and parents and staff, but less so in the language-reasoning subscale. Finally, the programme structure subscale has items with high, average, and low scores. It is worth noting that the item in this subscale on provision for children with disabilities could not be scored in as many as thirteen of the centres visited, because of the very limited enrolment in Early Start of children with identified disabilities. While the low participation of this group may reflect parental circumstances or choice as much as school policy, it is not consistent with the stated objectives of Early Start.

Individual Centre Scores

An examination of the mean scale scores of individual centres (calculated by summing the item scores for each subscale and dividing the total subscale score by the number of items scored) indicates differences between the centres with values ranging from 3.18 to 5.71 (Table 2). The total item score in the top-scoring centre is nearly twice that in the centre with the lowest score. Otherwise, the extent of variation may be illustrated in the finding

that eight centres achieved a mean scale score of at least 5, seven achieved a mean scale score between 4 and 5, while in the remaining four centres, the mean scale score is less than 4.

Table 2: Item Scores and Mean Scale Scores, by Centre

School	Total item Score	Mean Scale Score
1	160	5.71
2	157	5.60
3	154	5.50
4*	151	5.59
5*	148	5.48
6*	146	5.41
7*	142	5.26
8*	137	5.07
9	129	4.61
10*	124	4.59
11	121	4.32
12*	118	4.37
13*	115	4.25
14*	110	4.07
15*	108	4.00
16*	104	3.85
17*	100	3.57
18*	95	3.52
19*	86	3.18

* Mean scale scores are calculated excluding the item on disability

The subscales on which centres differed most are programme structure which has mean subscale scores ranging from 1.50 to 7.00 and personal care routines which has mean subscale scores ranging from 2.00 to 7.00 (Table 3). Somewhat smaller differences between centres are evident on the subscales for language-reasoning, interaction, space and furnishing, and activities. Between-centre differences are smallest on the parents and staff subscale which has values ranging from 2.00 to 4.66.

Table 3: Mean Scale Scores (Minimum and Maximum Values), by Subscale (N=19)

Subscale	Minimum Value	Maximum Value
Space and furnishings	2.75	6.50
Personal care routines	2.00	7.00
Language-reasoning	3.00	7.00
Activities	2.50	6.00
Interaction	3.25	7.00
Programme structure	1.50	7.00
Parents and staff	2.00	4.66

General Observations

In this section, the results of analysis of the observation schedule are discussed in light of some general observations of Early Start centres and interviews with staff.

The observation visits drew attention to the positive interaction between adults and children that featured in all of the centres visited. High standards of general supervision were maintained throughout, while only minor discipline problems were observed. Staff interacted gently with children, encouraging their efforts and achievements and supporting the development of interpersonal skills (e.g., turn taking and the promotion of fair play). Daily routines were enlivened with a mix of whole-group and small-group activities.

The physical environment of Early Start classrooms received a good deal of attention in the observation schedule and, confirming earlier evaluation findings (Educational Research Centre, 1998), it is clear that these aspects of provision are of a high standard. Rooms are spacious, bright and cheerful and there is an abundance of equipment and materials.

Analysis of the observation schedule results highlighted the focus in Early Start on language development. The visits to classrooms confirmed that provision of books and other language development materials for children (and parents) is excellent overall. Wall displays and room layout suggested that children were introduced to a wide range of topics including animals and nature, as well as domestic and work-related themes. Story telling is a major activity in Early Start and, in many classrooms, children were afforded some opportunity to make a contribution. Even so, there was relatively little one-to-one sustained interaction between children and adults as recommended in the Early Start dialogue strategies, and it would seem that more might be done to encourage individual children to complete full sentences and/or to express themselves more explicitly.

Observation of creative endeavours involving music, art, and socio-dramatic play proved somewhat difficult and yielded very little information about the standards achieved in these areas. It is not yet clear why this occurred. It may be because teachers and childcare workers were reluctant to engage in creative activities in the presence of observers, or that the children knew more than they were prepared to reveal, or simply that their teachers did not give them sufficient scope to demonstrate the extent of their knowledge during the visits. Some further investigation will be required before these questions can be addressed.

Cognitive development was supported by different kinds of table-top activity based on a variety of sorting materials (e.g., jigsaws, blocks, and beads) which children were encouraged to work with. The focus of adult-child interactions, however, tended to be on naming objects and some of their properties. There was little emphasis on number or on the attributes of materials other than size, shape, and colour and it would seem that the scope of Early Start was also not fully reflected in this domain.

Personal care routines (greeting/departing, meals/snacks, and safety code), an integral part of school experience, were not addressed to any extent in the Early Start guidelines, but were included in the observation schedule. In all of the centres visited, issues of safety (indoor and outdoor) were treated seriously and it seemed that every precaution was taken to avoid accidents and to promote awareness of potential hazards. In contrast, poor hygiene standards in relation to washing of hands before food handling were observed to an extent that seems surprising, given the separate toilet and washing facilities provided in nearly all of the units. Greeting and departing arrangements were orderly, friendly and positive. In a few centres, however, none of the parents entered the classroom when collecting their children, which seemed at odds with other efforts to involve parents including the fact that in ten of the centres visited, at least one parent or grandparent (mother or grandmother) was present for at least twenty minutes of activity. Teachers also pointed to the involvement of parents in school outings and in reading and play assignments at home, and to the contributions of the HSCL Co-ordinators in encouraging parent contact.

Teachers and childcare workers appeared to work well together. Responsibilities seemed clearly defined and, in general, the work seemed evenly divided. In some centres, a very positive air of teamwork was discernible; the staff acknowledged the supportive role of the school principal and reflected critically on their own progress.

Enrolment and attendance issues present an on-going challenge. A total of 267 children were enrolled in the centres visited. The maximum number that could have been enrolled, allowing for fifteen children per classroom, is 285. Most of the under-enrolment

is attributable to just two centres, however. In one of these, only ten children were enrolled and in the other only eight. Attendance problems are more widely experienced. During the visits, a total of 207 children were present: the numbers ranged from four to fourteen per classroom. Eleven or more children were present in thirteen classrooms but an additional three classrooms had just nine children, two others had only seven children, while in the remaining classroom there were just four children in attendance.

Conclusion

Application of the ECERS-R supports the conclusion that Early Start is a high quality intervention with a strong emphasis on adult-child interaction. Further, the information obtained in the study confirms previous evaluation findings that a number of improvements (e.g., greater parental involvement, better working-relationships between teachers and childcare workers, and a shift to small-group learning contexts) had occurred in the implementation of Early Start (Lewis and Archer, 2002). There were no indications from the observation data to contradict teachers' earlier reports of progress in these aspects of Early Start.

While the ECERS-R did reveal differences between centres, the instrument was not as discriminating as anticipated. Performance was superior on some subscales (interaction, language-reasoning, and space and furnishings) than on others, by and large reflecting the priorities given to classroom resources and to the language component of the programme. On the subscales on which centres did less well (personal care routines, parents and staff, and programme structure), most of the under-performance can be explained by low scores in a few items. In drawing attention to the low rates of enrolment of children with disabilities, the ECERS-R proved useful as a means of highlighting the mismatch between practice and policy. However, the appropriateness of the ECERS-R to capture and record progress in some of the more creative aspects of the curriculum must be called into question and some other research method explored.

A further limitation of the ECERS-R is the absence of provision for information on several key aspects of Early Start including assessment profiling, the teacher-childcare worker relationship, and the rich complexity of home-school-community liaison that supports Early Start beyond the immediate environs of the classroom. Further extensions of the scale would be required to ensure that it adequately reflects these major objectives of Early Start, a resource decision that may not be warranted in view of the scale of the programme taking into account its duration and intensity, as well as the number of centres involved. In conclusion, the ECERS-R provided valuable insights on the implementation of Early Start and, importantly, confirmed the views of teachers reported in questionnaires. Its contribution to Early Start is probably best appreciated as part of an array of research strategies deployed in the evaluation.

The authors are grateful to Séamas Ó hUallacháin who undertook some of the evaluation fieldwork.

References

Abbott-Shim, M. and Sibley, A. (1992). *Research Version of the Assessment Profile for Childhood Programs*. Atlanta, GA: Quality Assist, Inc.

Arnett, J. (1989). Caregivers in Day-care Centers: Does Training Matter? *Journal of Applied Developmental Psychology*, Volume 10, pp. 541-552.

Bredekamp, S. (1986). The Reliability and Validity of the Early Childhood Classroom Observation Scale for Accrediting Early Childhood Programs. *Early Childhood Research Quarterly*, Volume 1, pp. 103-118.

Educational Research Centre (1998). *Early Start Preschool Programme: Final Evaluation Report*. Dublin: Educational Research Centre.

Essa, E. and Burnham, M. (2001). Child Care Quality: A Model for Examining Relevant Variables (in) Reifel, S. and Brown, M. (Eds.). *Advances in Early Education and Day Care, Volume 11, Early Education and Care, and Reconceptualizing Play*. Oxford: Elsevier Science, pp. 59-113.

File, N. and Kontos, S. (1993). The Relationship of Program Quality to Children's Play in Integrated Early Intervention Settings. *Topics in Early Childhood Special Education*, Volume 13, pp. 1-18.

Harms, T. and Clifford, R. (1980). *Early Childhood Environmental Rating Scale*. New York: Teachers College Press, Columbia University.

Harms, T. Clifford, R. and Cryer, D. (1998). *Early Childhood Environment Rating Scale – Revised Edition*. New York: Teachers College Press, Columbia University.

Howes, C. and Stewart, P. (1987). Child's Play with Adults, Toys and Peers: An Examination of Family and Child Care Influences. *Developmental Psychology*, Volume 23, pp. 423-430.

In-Career Development Team in Collaboration with the Early Start Personnel (1998). *Early Start Pre-school Intervention Project: Curricular Guidelines for Good Practice*. Dublin: Department of Education and Science.

Kelly, D. and Kellaghan, T. (1999). *The Literacy and Numeracy Achievements of the First Cohort of Early Start Children (1994/95) when they were in Second Class (1998/99)*. Dublin: Educational Research Centre.

Lewis, M. and Archer, P. (2002). *Further Evaluation of Early Start. Progress Report*. Dublin: Educational Research Centre.

Lewis, M. and Archer, P. (2003). *Early Start Evaluation: Report on Observation Visits to Schools*. Dublin: Educational Research Centre.

McCartney, K., Scarr, S., Rocheleau, A., Phillips, D., Abbott-Shim, M., Eisenberg, M., Keefe, N., Rosenthal, S. and Ruh, J. (1997). Teacher-child Interaction and Child-care Auspices as Predictors of Social Outcomes in Infants, Toddlers, and Preschoolers. *Merrill Palmer Quarterly*, Volume 43, pp. 426-449.

Sylva, K., Siraj-Blatchford, I. and Taggart, B. (2003). *Assessing Quality in the Early Years. Early Childhood Environmental Rating Scale Extension (ECERS-E): Four Curricular Subscales*. Stoke on Trent: Trentham Books.