Weaknesses and strengths in assessing early childhood programmes: an assessment of an early childhood Spanish trilingual programme in two- to three-year-old children

Sonia Rivas, Angel Sobrino and Felisa Peralta

Department of Education, University of Navarra, Pamplona, Spain received 14 January; final version received 28 May 2008)

This article gives an account of the results from an assessment of an early childhood education programme, conducted over the course of two academic years (1999–2000 and 2000–2001), in a centre in northeastern Spain. The purpose of the assessment was to discover how a particular educational programme contributed to the short-term competency levels of children aged from two to three years old. The programme's curriculum encompassed different areas of development, including physical exercise and motor movement and social and linguistic development, using a unique teaching methodology that exposes the children to three different languages at the same time. The article includes a discussion of the weaknesses and strengths in implementing evaluation in early childhood education programmes, and concludes with some guiding principles that may prove useful in the evaluation of the appropriateness of any given assessment method used within an early childhood educational programme.

Keywords: early childhood education programmes; assessment; quality; early childhood education; standards; language immersion

State of the problem

Different fields, such as developmental psychology and neuroscience, as well as child psychology, concur that during the first years of life, activity which stimulates a child's central nervous system (CNS) has the potential to enhance his/her abilities and talents (De Graaf-Peters & Hadders-Algra, 2006; Lagercrantz, Hanson, Evrard, & Rodeck, 2002; Volpe, 2001). Several studies carried out in the field of special education demonstrate positive outcomes in the recovery of learning abilities, as a result of early intervention. At the moment, researchers tend to favour a cautious approach to programme design (Bruer, 2000; David, 1990; Little, 2001; Rivas, 2004; Schiller, 2001; Shelden & Rush, 2001; Stover, 2001) because the development and assessment of learning abilities in children is very complex (Shonkoff & Phillips, 2000; Slavenas, 1993). Nevertheless, it seems premature to limit child educational programmes that may enrich children's full potential because it has not yet been proven that certain forms of educational intervention provided during the first three years partially or wholly determine future ability and performance and academic achievement.

Acknowledgment of the fact that children under three years old are capable learners is reflected in Spanish legislation which recognises the educational nature of the interventions during the primer ciclo de Educación infantil, from ages 0 to 3, and requires pre-kindergarten teachers to offer a pedagogical programme (Ley Orgánica de Educatión [LOE; Organic Law of Education], 2006, Title I, Article 14, No. 3) and to guarantee quality practices (Ministerio de Educación y Ciencia [MEC; Ministry of Education and Science of Spain], 2006). Furthermore, the legal framework establishes a common curriculum for children from ages zero to six (Real Decreto 114, 23 January 2004), although the different comunidades autónomas in Spain have refined the curriculum in line with their specific forms of competence and management. At the same time, individual schools have designed and organised their own education programme, respecting the common curriculum and the educational nature of this period. As a result, remarkable variety in early childhood educational (ECE) services has emerged. Although research and legislation ought to guide this transition, many of the programmes developed and implemented in Spain are not based on researchers' recommendations (Aguado & Jimenez, 1998) or on a realistic vision of the first three years of childhood (Rivas, Sobrino, & Peralta, 2005). At the moment, teachers follow their own criteria in deciding if a programme has fulfilled the legal requirements concerning quality.

There is an increasing interest among researchers to understand how quality is established and maintained in ECE programmes (Clark & Stroud, 2002; Cryer, 1999; Fontaine, Torre, Grafwallner, & Underhill, 2006; Gol-Guven, 2007; Koralek, Colker, & Dodge, 1995; Lee & Walsh, 2004; Sarancho & Spodek, 2007; Sheridan & Schuster, 2001). However, policies concerning the development of ECE programmes occasionally include guidelines or protocols which draw on the quality principle. Spanish legislation does not define how assessment be carried out or what quality means during this early period. As a result, although the definition of quality is important, how this quality is to be produced and judged in childhood programmes, apart from what factors play a role in quality, what constitutes a quality programme, how a quality service is to be realised in practical terms, how different levels of quality may be evaluated and how quality data may be interpreted must also be underlined.

Consequently, due to the fact that ECE programmes are exceedingly diverse, it is essential that consistent ways of judging the value of programmes be developed and the best structure for ECE programmes be defined.

The present study

A programme evaluation was designed and conducted over the course of two academic years (1999–2000 and 2000–2001), to study a particular programme's contribution to children's short-term competency levels. If this particular programme is shown to improve children's performance, we may also have taken a small step forward in determining how ECE programme can be assessed from an empirical point of view.

Assessment data came from different stakeholders: children, teachers and setting1 although not all of them will be described. Furthermore, the evaluation model developed by Perez Juste (2000), as well as the Glasman and Nevo (1988) model for decisionmaking, was used as a foundation for the assessment protocol applied in this study.

Background

Centre description

The programme under consideration is called the Trilingual Early Stimulation Programme (Programa de Estimulación temprana y trilingüismo), and is implemented in a private full-service childcare and education centre located in the Basque Country region, in north-eastern Spain. The centre is located in a detached house and includes Catholic catechism instruction and meal times as part of its service. The facility provides full-time and part-time care and education to children from four months to three years of age. In general, this programme operates five days per week, for an average of seven and a half hours per day. The centre also offers educational and care services on Saturday mornings. This centre has been in existence for 10 years and serves over 50 families, most of whom belong to a professional, middle to upper socioeconomic class.

The nine staff members at this ECE centre (100% white female) come from a variety of educational backgrounds, with qualifications in education and early child-hood training and fluency in different languages (Cambridge Certificate in Advanced English [CAE] and Proficiency level in English, Degree level in Spanish and Euskararen Gaitasun Agiria [EGA; Official Certificate in knowledge of Euskera] level in Basque). The age range of full-time staff members is between 23 and 33 years old. The length of time of service for staff members ranges from two to nine years. In each classroom, there are 11 children and two teachers, independent of the activity of the children in the group, the age of the children and the legally suggested staff–child ratio (1:8).

Programme description

The programme curriculum has a strong academic approach and focuses on different areas of development, including physical exercises and motor, social and language skills, using a unique teaching methodology characterised by language immersion. In line with the trend and expectations of childhood bilingualism that have emerged in recent years (Baker, 2006; Chin & Wigglesworth, 2007; Creese, 2005; McCardle & Hoff, 2006), and specifically in the Basque Country (Sierra, 2008; Zalbide & Cenoz, 2008), the staff use this expression to evoke the image of submerging children in an environment where everything that surrounds them foments the learning of three different languages.

The methodology of this programme is based on teacher rotation. Each staff member teaches – in Spanish, English or Basque – for 20 minutes in every class, irrespective of the ages of the toddlers and infants. Language exposure occurs during different routines over the course of the day: arrival and departure, outdoor and indoor play, meals, toileting and educational activities. Thus, the children hear all three languages spoken by different teachers who rotate through their classrooms everyday and in the same routine situations; the two teachers in the classroom at any one time also have different roles: one is the leader, the other the assistant. In relation to classroom management, the teachers adopt a relatively uniform approach to strictness.

Activities outlined in the curriculum include structured gross motor skills exercises (crawling, jumping and rolling) and structured fine motor skills exercises with rhythmic accompaniment, as well as emergent literacy development activities (see Table 1). The programme curriculum also addresses cognitive areas such as science, mathematics, geography and art, using materials like flashcards (called bits de inteligencia) for teaching and methodologies like musical hearing, cultural walks and directed group-play activities.

The structured and sequenced academic tasks are designed to introduce children to facts and knowledge that they are unlikely to learn spontaneously or by discovery, such as the name and recognition of Velazquez's paintings or the Latin names of all kinds of pine trees. These tasks involve memorising lists and pictures, responding to questions with correct answers and practising routine tasks that can be assessed as right or wrong. The primary activities centre on directive instruction, and systematic work is conducted on oral language and emergent literacy skills. When curricular content is covered, it is normally carried out through a teacher-centred, whole-group format.

Table 1. Time distribution in the schedule at the school.

		Physical area		Teaching strategies	
Schedule		Indoor	Outdoor	Structured	Unstructured
8:00–9:00	Children's arrival and welcome	→		→	
9:00–9:30	Indoor, directed play	→		→	
9:30–9:50	Cognition area: Basque language exercises	→		→	
9:50–10:10	Cognition area: Spanish language exercises	→		→	
10:10–10:30	Cognition area: English language exercises	→		→	
10:30–10:50	Fine motor and audition exercises	→		→	
10:50–11:15	Gross motor exercises and clearing time	→		→	→
11:15–11:45	Outdoor play		→		→
11:45–12:30	Mealtime	→		→	
12:30–13:15	Nap	→		→	
13:15–15:45	Indoor, directed activities and mealtime	→		→	
15:45–17:00	Nap and children's departure	→		→	→

According to the programme policy, carrying out these activities stimulates the development of the children's CNS. It maintains that children may acquire a variety of skills during these three years of life because they are especially capable of learning during this period (Doman, Doman, & Aisen, 2005).

Consequently, it holds that children ought to be immersed in stimulating environments that enable their learning, always following 'the law of ability' methodology which shows children a

model and provides them with the opportunity to copy it in order to facilitate the acquisition of certain skills. Furthermore, the individualisation in learning principle is followed: in so far as possible, the programme adapts to and respects individual differences and the unique emotional development of each child.

The programme encourages parental involvement and emphasises the importance of linking teachers and families because the teachers subscribe to the concept of shared educational responsibility, which implies recognition of parents as the primary educators of children. Thus, the teachers work closely with parents to create and maintain contact through direct and indirect channels of communication.

Teachers encourage the children's emotional growth and learning in class, and the children's identification of what is desirable and what is not, by using a social reward system: a sticker for the child's clothes; a sticker is also placed in the personal diary so that parents can reinforce particular behaviours in the child: the case, for example, that a child tells the teacher about going to the bathroom when he/she is being toilet-trained.

Research question

The question that shaped the evaluation was to determine if this programme contributed to the development of the children's abilities and skills in the motor, language, cognitive and social areas, as part of a wider assessment of the programme (Rivas & Sobrino, in press). To accomplish this objective, a specific protocol was designed and applied in this case study (Rivas, 2008).

Information was gathered from two sources – teachers and setting – to garner an in-depth understanding of the meaning of the development in the children's learning abilities. The quality of the setting at the centre as an effective context for the development of children's abilities was considered. An association between setting and programme quality may be established on the ground that the distribution of the learning area may modify the children's interests, their levels of interaction and their relationships with one another (Fernandez & Gonzalez, 1997; Schwartz & Olswang, 1996).

Variables

Seven specific variables and a total of 270 indicators drawn from the specification of four dimensions in the Haizea instrument (Fuentes-Biggi, Fernandez, & Alvarez, 1992) were taken into account. Furthermore, three of the seven subscales of the Spanish version (SV) of the Early Childhood Environment Rating Scale (ECERS; Harms & Clifford, 1980), adapted by Palacios and Lera (1992), were included in the study.

Methodology and design Sample size

This research was undertaken when the programme had already been established and was in process, with a specific number of children attending every class in the programme. We gathered data for 11 of the children in the research sample, at or near the age of two,2 who

were selected according to two requirements: (1) each child was going to enter the programme and was at least two years old, the age that corresponds to the application baseline of the Haizea instrument; and (2) each child was going to remain in the same city for the two academic years.

Design

Both quantitative and qualitative methodologies were applied in the present research. The quantitative methodology was used to collect and to analyse information about the children. T-scores were employed (Mean = 50, Standard Deviation = 10) as a measure of variability. The request for a research project to assess this programme came when it was already in progress and with a specific number of children. Therefore, it was not possible to assign children to different groups or to have a control group with children of the same profile. The lack of control group was adjusted a priori because the instrument that we used to assess the children's abilities and skills was standardised for a very similar population. As a result, the sample assessed using this instrument served as a control group in this research.

Instruments

Two instruments were used in quantitative assessment to collect information: the Haizea instrument (Fuentes-Biggi et al., 1992) and the SV of the ECERS scale (Harms & Clifford, 1980), adapted by Palacios and Lera (1992).

The Haizea instrument (Fuentes-Biggi et al., 1992) allows children's maturation between birth and five years of age to be checked, and forewarns about autism and other disabilities. It is composed of 270 items that analyse the social, cognitive, motor and language skill areas, and five further sub-areas: self-care, fine motor, gross motor and expressive and cognitive language skills. Each skill includes specific behavioural components that are presented as performance criteria. In general, these behaviours represent a mature pattern of the skill. If the child performs a behavioural component correctly, the assessor marks a '1'; if the child does not perform a behavioural component correctly, the assessor marks a '0'.

Haizea instrument (Fuentes-Biggi et al., 1992) has an established reputation in Spain, and is regarded among early intervention professionals to be an excellent instrument for screening (Iceta & Yoldi, 2002; Llanos & Azurmendi, 2002). Essentially, it takes into account the questions set out in this study, especially those areas in which the programme has special interest:

(1) To evaluate the development of children in different regions, it is preferable to use a test with certain reliability to the reference population. The Haizea instrument was composed using data collected from 817 children in the Basque Country (with ages between two- and five-year-old) (Fuentes-Biggi, Fernandez, & Alvarez, 1991), the community where the research was carried out.

(2) It enabled us to establish global scores in different areas for children between the ages of two and five, the same areas for which this programme has been developed and the range in which the centre's principal has greatest interest.

(3) The Haizea instrument had more items per area than other similar instruments. Furthermore, it was standardised at monthly intervals, a fact that would facilitate later interpretation through the collection of more information in each area.

(4) Its statistical study (Fuentes-Biggi et al., 1991) guarantees internal validity (Reliability 0.84).

(5) The use of this test with children did not require more training than the evaluator already had.

(6) It offered the possibility of analysing the results in T-scores, the best way to provide a descriptive analysis, as had been planned.

The ECERS (Harms & Clifford, 1980) has been widely used in child-care research. Given that an SV of the original ECERS instrument (Harms & Clifford, 1980) is available (Palacios & Lera, 1992), it was used to gather information about the quality of the centre setting. This instrument is a measure used for research and programme evaluation purposes to assess the overall quality of classrooms in the Spanish context. The scale supplies a global quality score for early childhood education classrooms and facilitates an assessment of the developmental appropriateness of classroom practices by assessing routine care needs, furnishing and display, activities and experiences related to motor, language, cognitive and social development and provisions for adults. The SV of the ECERS contains seven subscales with a total of 37 items. It uses a sevenpoint system of scoring, with a score of '1' indicating inadequate, '3' minimal, '5' good and '7' excellent. In this research, three subscales were also used.

Data collection

The research process was characterised by flexibility, linearity and uninterrupted continuity. Children's scores were collected during the academic years 1999–2000 and 2000–2001. Data collection was completed by a main researcher at three different stages, which prevented the problems often associated with comparing information gathered by different observers. The language used to measure the receptive and expressive sub-areas was Spanish, because it was the main researcher's first language, although all children were early-stage bilingual (Basque and Spanish).

Stage 1

The first assessment was conducted on the children's initial entrance to the programme in the academic year 1999–2000, and gathered information on children enrolled in the programme. During that period, the researcher applied the Haizea instrument between the months of April and June 2000 in order to collect information which reflects the development of the children in different learning ability and skill areas.

All children were tested individually, in the multi-purpose room at the kindergarten school, on the children's initial entrance to the programme and after the implementation of the

intervention programmes. Prior to the test, the researcher provided standardised verbal instructions to each child. The children were encouraged to make their best effort.

The quality of setting area report was gathered using the SV of ECERS (Palacios & Lera, 2002) at the start of the second year. The main observer completed the SV of ECERS in each classroom during a 3- to 4-hour observation period, although each classroom with an assistant teacher was observed constantly over the course of the two academic years. The main observer visited classrooms independently and completed other background information on teachers, group size and teacher/child ratios. The children, teachers and settings were observed during toileting, meals, and play and afternoon activities from Tuesday to Friday.

Information about interaction between children and teacher and teacher behaviour was collected using a checklist. Positive comments from teachers and children were checked in five different situations and in a variety of different circumstances (Rivas & Sobrino, In press).

It should be noted that the limitations of more traditional forms of assessment of young children were ameliorated to the best of our ability. The lack of motivation which may occur in children's performances due to uninteresting and/or unfamiliar material was avoided. All the elements in the Haizea instrument were attractive and engaging for the children. Furthermore, one expert observer in this research project worked with the group of children for the entire academic year, so the failure of children to perform because of an unfamiliar situation or an unknown staff member was likewise avoided.

Stage 2

The researcher conducted the second assessment using the Haizea instrument between the months of April and June 2001, one year later, with different averages for the ages of the two applications. Many children had moved to other early childhood settings, so the data collection had to be carried out in different physical environments.

Stage 3

During the third stage, the information was analysed and the analysis focused on how to create a proposal for decision-making to improve this programme, as well as on how to inform the centre's principal about the results.

Results

In this section, the results obtained from the children in the different areas and subareas will be briefly outlined. These results show the scores from the first and second assessments (see Table 2). Nevertheless, the gains that these children made imply the short-term positive impact of educational intervention on the cognitive development of children who do not have cognitive difficulties. The table comprises data on 11 kindergarten children, on the children's initial entrance to the programme (first application) and after the implementation of the intervention programmes (second application). While these results suggest that these children are growing in their emergent and early literacy skills, they do not appear to suggest significant development in other areas.

Table 2. T-score distribution in social, cognitive, motor and language areas in both applications.

					Motor area				Language area			
	Social area		Cognitive area		Fine motor		Gross motor		Receptive		Expressive	
No.	Application 1	Application 2	Application 1	Application 2	Application 1	Application 2	Application 1	Application 2	Application 1	Application 2	Application 1	Application 2
1	60	57	52	53	47	59	51	40	33	48	33	48
2	69	46	58	54	57	41	39	61	50	45	50	45
3	56	50	57	55	56	44	50	58	59	36	59	36
4	43	33	36	38	43	3	27	29	12	49	12	49
5	63	55	50	53	44	55	24	55	53	56	53	56
6	55	56	60	59	58	58	52	58	48	51	48	51
7	51	54	48	60	43	54	37	60	39	59	39	59
8	40	65	46	59	50	57	33	58	31	56	31	56
9	38	54	42	47	37	54	37	38	33	49	33	49
10	35	50	54	51	31	54	37	51	33	59	33	59
11	57	42	47	36	46	60	40	31	33	57	33	57

T-scores in the general language area show that nine of the students improved on the scores in the first application, while two did not. If the comprehensive or receptive and expressive sub-areas are analysed separately, the data implies that the children had higher scores in the expressive sub-area than those for the normative group in the Haizea instrument. However, children obtained scores similar to those for the normative sample in the receptive language sub-area. In the cognitive area, the results seem to indicate that this programme is effective in the development of abilities measured in the Haizea instrument for this area: form, magnitude, the spatiotemporal dimension, the memory line of vision-perception-discrimination and numbers. The assessment results may be considered satisfactory because 8 of the 11 children in the sample placed in the average range (between the average and +1 SD), while three children placed between the average and ❑1 SD. The Haizea instrument supplies general results for the motor area, apart from the results for different sub-areas: fine motor and gross motor skills. T-scores in the fine motor sub-area reveal that eight of the children in the group improved their scores, whereas three disimproved. In addition, T-scores in the gross

motor sub-area show that nine of the students improved their results in the second application of Haizea instrument, whereas two did not.

Different abilities such as physical self-knowledge, ability to play, interpersonal relationships and personal autonomy (sphincter control, toileting, habits and skills in dressing and during meals) is assessed in the area of socialisation. Specifically, less progress was noted in this area in comparison to other areas assessed in the study because only 6 of the 11 children in the sample placed in the average range (between the average and +1 SD), while 5 children placed below (between the average and ❑1 SD).

Limitations, unanswered questions and issues

As much as knowledge has advanced about the effects of ECE evaluation programmes, many questions remain unanswered: What are the causal mechanisms and pathways through which the effects of a diversity of early childhood intervention strategies promote long-term success? For whom are existing programmes most effective, and which programme features are most associated with success?

These and other questions represent, in part, the natural progression of research towards a more complete cumulative knowledge base. Conversely, related questions concerning which programme components are most closely tied to children's success are subject to significant limitations, which are widely acknowledged and difficult to resolve: the programme outcomes that yield the largest as well as the smallest effects; the size of the programme effect; the mechanism through which the estimated effects are manifested; the factors that may influence the programme; or the use of small samples. This particular study was subject to a number of limitations:

(1) The first limitation refers to the determination of variables within ECE programmes. The majority of assessments currently used within ECE settings are designed to provide a great deal of information about a child's development through variables (and, indeed, often an unspecified number of variables) that are not under teacher control, and that are vastly greater than the number of children in the experimental samples. Thus, the attribution of changes to programme impact needs to be approached with caution. Evaluation methodology with ECE programmes remains significantly linked to the tradition of

Campbell and Stanley (1966) and Cook and Campbell (1979), in which main effects and internal validity are emphasised over issues concerning moderators and mediators, generalisability and programme features associated with issues. As a result, the possible determining variables that could account for observed changes in children's competence within ECE programmes, and within this specific programme, are large in number and difficult to control and measure. Many such assessments are considered valid because they have demonstrated some level of predictive validity in psychometric analyses. However, good predictive validity should not make a variable a legitimate target for assessment in an ECE classroom; legitimacy in the classroom context should be read in terms of intervention, not prediction (Downs & Strand, 2006). The position adopted in this project was that the value of any assessment is dependent

on the action that it affords the person to whom the information is directed. As such, what constitutes valuable information for educators may often be different to what constitutes valuable information for policy-makers, administrators or psychometricians.

(2) The second limitation is in relation to causal uncertainty – that is, the determination of whether the observed gains (which may be similar to the gains observed in other situations) are due to this programme's effects or participation in other activities or normal child development. Literature in this field highlights a specific criterion that is not fail-safe but which helps build a case for interpreting the effects of ECE programme when experiments are not possible (Reynolds, 2004). According to the gradient criterion, a causal inference is more warranted if, all other factors being equal, a monotonic relationship exists between programme exposures (number of days or sessions attended, number of contact hours, number of years of participation) and the programme outcome.

(3) The third limitation concerns the mechanism through which the estimated effects are manifested. All assessments of young pre-school children are notoriously unstable, which means that the standard errors of various tests used may be large. In this regard, it should be noted that the limitations of more traditional forms of assessment of young children were ameliorated to the best of the researchers' ability in this project. The lack of motivation which may occur in children's performances due to the use of possibly uninteresting and/ or unfamiliar material was avoided. All the elements in the Haizea instrument were attractive and engaging for the children. Furthermore, one expert observer involved in this research project worked with the group of children for the entire academic year, so the failure of children to perform because of an unfamiliar situation or an unknown staff member was avoided. Perhaps the most serious limitations on this study (at the time the project was carried out) were the absence of data regarding the intensity of services delivered and the lack of outcome data to assess the effectiveness of these services for children.

(4) The fourth limitation is in relation to generalisability. Studies of the comparative effects of programme curricula have been limited mostly to children who have learning difficulties. Furthermore, specificity of association refers to the situation in which the programme–outcome relationship is limited to certain domains of behaviour or to particular sub-groups such as individuals with disabilities. Causal inference is more straightforward in such cases. In addition, evaluations are frequently conducted in a comparative design format, for example, a contrast between those who have been given the programme and those who have not. Because the manner of assignment of schools to programme is almost exclusively non-random, and participation is based on the high motivation of school-community staff, causal relationships are strictly not inferable. The alternative hypothesis for any programme effects is rooted, of course, in the potential pre-programme differences in the participating schools. The present study sample comprised Basque Country children – living in this city. Therefore, the findings should not be broadly generalised to other populations or other contexts, although the study may support findings from other early childhood programmes conducted in other locations. Thus, further replications are needed to establish the generalisability of the results.

Summary and implications

The universal objective for teachers and kindergarten schools in Spain is to promote the overall development of each child. However, from now on, teachers are also required to design and implement a high quality educational programme for each child. Given the increasing number of kindergarten schools and teacher autonomy in designing curricula, a serious effort must be made to understand what constitutes the best curricula for children and what activities should be included, especially because of the growing insistence among parents that children from birth to age three participate in learning programmes with taught curricula, using methodologies that focus on the acquisition of language and motor skills.

This centre in the Basque Country, which emphasises an academic approach in different areas of development (motor, language, cognitive and social) for children from birth to age three, was assessed using a unique methodology to see if a particular programme based on this perspective really helps children to acquire the targeted abilities.

The major finding of this study is that the programme, its particular teaching methodology and environment, contributes in a moderate way to short-term intellectual, social, motor and language development in children from two to three years old, between Stage 1 (between April and June 2000 for the Haizea instrument) and Stage 2 (between April and June 2001 for the Haizea instrument).

Readiness means that learning is appropriate to the child's development as well as to the child's chronological age. In this sense, educational programmes for infants should not become product-oriented, 'in order to please adults and prepare them for the competitive, test-driven culture that awaits' (Jalongo et al., 2004, p. 145). Thus, time management and methods of teaching in pre-school education must be founded on a respect for early childhood as a unique period, rather than a drive to replicate the curriculum and pedagogy that characterises later academic experiences.

Arriving at a definition of what constitutes a high quality educational programme for very young children remains a challenge. As Jalongo et al. argue:

Curriculum guides should not become absolutes; rather, they should be viewed as works in progress as they are continuously improved to reflect new knowledge about how young children learn and are adapted to better serve particular groups of children and families. (2004, p. 145). Consequently, more research is needed to address these questions because knowledge of what practices pre-school programmes should embrace may broaden the current understanding of best practices in ECE and influence and improve the current system of educational programmes for children from birth to three years of age.

In an effort to respond to what researchers and policy-makers report about the need to document developmentally appropriate models for assessing ECE programmes, some practical implications are mentioned here. It is important to point out that the diversity found in the international community, due to different political systems, cultural differences and dissimilar societal commitment to validating wider assessment in ECE (Guralnick, 2008), is a significant barrier to the establishment of a consensus. The presentation of the following information seeks to minimise bias and maximise agreement in developing universal practices and procedures for ECE assessment.

Use and content of assessment

An authentic approach to the assessment of educational programmes for very young children ought to meet many of the defining purposes of such assessment: (1) to identify children who are educationally at-risk, and in need of specialised intervention and educational services (Bagnato, 2005; Macy, Bricker, & Squires, 2005); (2) to provide information on child learning and progress; (3) to evaluate trends in service utilisation, provision and quality for policy-makers and planners (Kagan, 2003); and (4) to foment accountability in effective assessment provision (Rous, Lobianco, Moffett, & Lund, 2005).

The aim of the purposes detailed above is to provide educators with information that enables the maximisation of intervention effectiveness in relation to each stated objective. In other words, the use of assessment should be measured in terms of its impact on decision-making, in accordance with final results.

Assessment procedure

In order to accomplish the uses mentioned above, an assessment is to be regarded as a sequence of events. Firstly, at the beginning, process variables should be measured. Thus, assessment may be used to guide curricular and instructional decisions to provide educators with information in a continuous formative process of assessment and to provide ongoing feedback to educators regarding how well everything is going in the programme, thus allowing educational modifications to be made as necessary. Consequently, ECE programme assessment components should be associated consistently with smaller effects, which may include teaching parents problem-solving, to promote children's cognitive, academic and/or social skills and providing other, additional services. Such assessment is better able to document progress over time as it is related to the goals of the programme.

Secondly, at the end, outcomes should be taken into account, enabling the analysis of the relationship between means and ends. As a result, assessment would provide final information about programme effectiveness. Consequently, ECE programme assessment components should also be associated consistently with larger effects, which may include, for example, increasing positive parent–child interactions and emotional communication skills, teaching parents to use 'time out' and the importance of consistency in parenting practices and requiring parents to practise new skills with their children during parent training sessions.

Frequency of assessment

An adequate frequency of assessment, which takes the final and summative evaluation goals into account, is also a key element in an authentic assessment approach. Assessment should be carried out three times per year, in a summative way. Since such measures would be used

repeatedly over the course of the school year(s), they could provide clear demonstration of development (or lack of development) in the desired area.

Additionally, bi- or tri-annual assessments would provide educators with valuable information at those time points in a conclusive way. Thus, by attending to summative and final evaluation goals, repeated performance assessment would enable educators to measure child development in relation to desired outcomes at the beginning of intervention and over time, over the course of almost two years.

Sources of assessment

An authentic assessment approach should be based on multiple sources of assessment, which reflect the complexity of the data involved. Hence, the importance of using parent–teacher collaboration to collect information such as an individual child's skills is emphasised, while information about functional behaviours relating to family and school stimuli in different natural settings is also engaged. Consequently, the approach promotes a high degree of family participation, and should draw on the commitment of dedicated staff.

Furthermore, assessment should be based on the documentation of group projects and collaborative work about teacher development guidelines and checklists; and, finally, on the use of summary reports (such as the personal diary) to provide detailed assessment of child-specific sources in certain domains.

Requirements for research design

The significance of different requirements in research design ought to be taken into account. Irrespective of the specific details of a given case, in order to meet the objectives of ECE programme assessment outlined above, ECE measurement and research design should be based on three major features: authenticity, utility and universality (Bagnato, 2005). This methodological framework for the assessment should combine most of the essential elements for recommended practices in assessment identified by the Division for Early Childhood (Neisworth & Bagnato, 2005). These elements should include (1) standardised tasks, but using flexible and graduated scoring and administration procedures; (2) natural observation assessment to respect functional assessment in the natural environment; and (3) adequate sample size.

Standardised tasks

The overlap of instruments used in the assessment of ECE programmes is remarkable, and demonstrates a widespread consensus regarding the definition of ECE throughout most of Europe. However, the instruments vary in scope and differ in detail. Some instruments evaluate ECE quality for all the children in a group, while others attempt to evaluate on an individual basis. Whatever the case, the usefulness of an assessment approach that reflects children's functional repertoires should be acknowledged. The main strength of the standardised testing approach is that such tests allow comparison with a normative sample, thus providing information regarding where an individual child's development stands in relation to his/her same-age peers. Norm-referenced assessments that yield standard scores (T-scores, Z-scores or deviation quotients) provide numerical scores that are interpretable

along a known distribution (normal curve).

Natural observation assessment

Assessment procedures that capture real-life competencies in everyday settings and document even small improvements in developmental skills should be a requirement in ECE programmes. Most everyday skills in daily routines are overlooked in the unnatural contrivances of testing situations and tasks. Consequently, assessment in natural observational assessment would have abalancing effect: it should be committed to hybrid measurement that is curriculum-based, but which is also an authentic assessment instrument in that the child's skills and behaviour are evaluated through the observation of children during real-life tasks.

Sample size

The numbers of possible variables operating in the experimental classroom are vastly greater than the number of children in the sample, so the attribution of changes to programme impact should be treated with great caution. It is important that the proportion of the total birth-through-age-2 population served in the ECE programme assessment be as large as possible – up to more than 30.

As a result of what has been discussed above, this fuller assessment would provide data that could be used to analyse results, utilisation and needs, and would allow families to see and understand their children's progress.

Furthermore, other aspects may emerge as contributing to the success of the assessment development process: establishing a shared vision of the goals of the ECE assessment programme and its outcomes; maintaining flexibility in implementing the phases and details of the programme; negotiating common understanding with participants and ensuring fruitful collaboration in planning and implementation. A great deal of work still remains to be done in this regard.

Acknowledgements

The authors wish to acknowledge and thank the anonymous reviewer of this article as well as Karen Murphy, PhD, from Wheelock College (Boston), and Cóilín O hAodha from the University of Navarra (Spain). Their thoughtful comments and suggestions on earlier drafts of the article are greatly appreciated.

Notes

1. Information came from the application of the Haizea instrument (Fuentes-Biggi, Fernandez, & Alvarez, 1992) with 11 children; from 9 interviews with centre staff and 11 with parents; from observations over different periods of time in classroom environments using ECERS-R (Harms et al., 1998), and from programme documentation.

2. Children number 4 and 5 were students with special needs.

Notes on contributors

Sonia Rivas is an Assistant Professor in the School of Humanities and Social Sciences at the University of Navarra. She has a PhD in Education from the University of Navarra. She lectures

in the Master's programme organised by the Institute of Family Sciences (MMF@) and the Institute of Religious Sciences (ISCR) of the University of Navarra. She has also been a Visiting Research Scholar at the University of Cambridge, UK, and at Wheelock College, Boston, MA. She is the author of a number of papers and articles on family and the improvement of quality in educational intervention.

Angel Sobrino is the Subdirector of the Department of Education at the University of Navarra. She has a PhD in Education and an MA in Psychology from the University of Navarra. She lectures on 'Individual Differences and Education' in University of Navarra. His field of interests include the evaluation of programmes and, specifically, those related to the 0–3 age group.

Felisa Peralta is the Director of the Department of Education at Navarra University. She has a PhD in Education from the University of Navarra. She has developed several courses, and has delivered papers and lectures on her area of research interest. She has also been Coordinator in Spain for a Transnational Socrates Project. She has published four books and written several chapters in collections, as well as articles relating to diagnosis and intervention in handicapped people with special needs. Her current research is centred on self-determination.

References

Aguado, T., & Jimenez, R. (1998). Early childhood education: Research in Spain. In T. David (Ed.), Researching early childhood education (pp. 111–130). London: Sage.

Bagnato, S.J. (2005). The authentic alternative for assessment in early intervention: An emerging evidence-based practice. Journal of Early Intervention, 28, 17–22.

Baker, C. (2006). Foundations of bilingual education and bilingualism (4th Rev. ed.). Clevedon, UK: Multilingual Matters.

Bruer, J.T. (2000). El mito de los tres primeros años. Una nueva visión del desarrollo inicial del cerebro y del aprendizaje a lo largo de la vida [The Myth of the first three years. A new understanding of early brain development and lifelong learning]. Barcelona: Paidós.

Campbell, D.T., & Stanley, J.C. (1966). Experimental and quasi-experimental design for research. Chicago, IL: Rand McNally.

Chin, N.B., & Wigglesworth, G. (2007). Bilingualism. Routledge: London.

Clark, P., & Stroud, J. (2002). Working together to improve the quality of care and education in early childhood programs. Early Child Development and Care, 172, 55–63.

Cook, T.D., & Campbell, D.T. (1979). Quasi-experimentation: Design & analysis issues for field settings. Boston: Houghton Mifflin.

Creese, A. (2005). Teacher collaboration and talk in multilingual classrooms. Clevedon, UK: Multilingual Matters.

Cryer, D. (1999). Defining and assessing early childhood program quality. Annals of the American Academy of Political and Social Sciences, 563(1), 39–55.

David, T. (1990). Under five, under-educated? Buckingham: Open University Press.

De Graaf-Peters, V.B., & Hadders-Algra, M. (2006). Ontogeny of the human central nervous system: What is happening when? Early Human Development, 82(4), 257–266.

Doman, G., Doman, J., & Aisen, S. (2005). How to give your baby encyclopaedic knowledge: More gentle revolution. Garden City Park, NY: Square One.

Downs, A., & Strand, P.S. (2006). Using assessment to improve the effectiveness of early childhood education. Journal of Child and Family Studies, 15, 671–680.

Fernandez, M.J., & Gonzalez, A. (1997). Desarrollo y situación actual de los estudios de eficacia escolar [The current situation and development of studies in school efficiency]. Revista Electrónica de Investigación y Evaluación Educativa, 3(1–3), 1–10.

Fontaine, N.S., Torre, L.D., Grafwallner, R., & Underhill, B. (2006). Increasing quality in early care and learning environments. Early Child Development and Care, 176(2), 157–169.

Fuentes-Biggi, J., Fernandez, I., & Alvarez, E. (1991). Tabla de desarrollo psicomotor [Psychomotor development chart]. In Estudio Haizea Llevant. Vitoria: Servicio central de publicaciones del Gobierno Vasco.

Fuentes-Biggi, J., Fernandez, I., & Alvarez, E. (1992). Escalas Haizea-Llevant para la evaluación del desarrollo de 0 a 6 años [The Haizea-Llevant scales for the evaluation of development in 0–6 year-olds]. Vitoria: Gobierno Vasco y Generalitat de Cataluña.

Glasman, N.S., & Nevo, D. (1988). Evaluation in decision-making: The case of school administration. Boston, MA: Kluwer.

Gol-Guven, M. (2007). Evaluation of the quality of early childhood classrooms in Turkey. Early Child Development and Care, DOI: 10.1080/03004430701217639

Guralnick, M.J. (2008). International perspectives on early intervention. A search for common ground. Journal of Early Intervention, 30(2), 90–101.

Harms, T., & Clifford, R.M. (1980). Early childhood environment rating scale. New York: Teachers College Press.

Harms, T., Clifford, R.M., & Cryer, D. (1998). Early childhood environment rating scale-revised. New York: Teachers College Press.

Iceta, A., & Yoldi, M.E. (2002). Desarrollo psicomotor del niño y su valoración en atención primaria [Child psychomotor development and its evaluation in primary care]. Anales, 25(Suppl. 2), 35–43.

Jalongo, M.R., Fennimore, B.S., Pattnaik, J., Laverick, D.M., Brewster, J., & Mukutu, M. (2004). Blended perspectives: A global vision for high quality early childhood education. Early Childhood Education Journal, 32, 143–155.

Kagan, S.L. (2003). Children's readiness for school: Issues in assessment. International Journal of Early Childhood, 35, 114–120.

Koralek, D.G., Colker, L.J., & Dodge, D. (1995). The what, why, and how of high-quality early childhood education: A guide for on-site supervision (Rev. ed.). Washington, DC: National Association for the Education of Young Children.

Lagercrantz, H., Hanson, H., Evrard, P., & Rodeck, C. (2002). The newborn brain. Neuro-science and clinical applications. Cambridge: Cambridge University Press.

Lee, J.H., & Walsh, D.J. (2004). Quality in early childhood programs: Reflections from program evaluation practices. American Journal of Evaluation, 25(3), 351–373.

Little, C.A. (2001). The myth of the first three years: A new understanding of early brain development and lifelong learning. Gifted Child Quarterly, 45(3), 226–227.

Llanos, M., & Azurmendi, M.J. (2002, October). Adquisición léxica del euskara desde un programa de inmersión total precoz: un estudio longitudinal (pp. 1091–1102). Vigo: Actas del II Simposio Internacional sobre Bilingüismo.

LOE. (2006). Ley Orgánica 2/2006, de 3 de mayo, de Educación (BOE May 4, 2006).

Macy, M., Bricker, D., & Squires, J. (2005). An examination of the validity and reliability of a curriculum-based assessment approach used to determine eligibility for Part C early intervention services. Journal of Early Intervention, 28, 295–310.

McCardle, P., & Hoff, E. (2006). Childhood bilingualism. Research on infancy through school age. Buffalo, NY: Multilingual Matters.

MEC. (2006). Evaluaciones generales del sistema educativo español. Apuntes del Instituto de Evaluación, no. 8 [Evaluation of the educational system in Spain. Notes from the Institute of Evaluation, No. 8]. Retrieved May 28, 2008, from http://www.institutodeevaluacion.mec.es/contenidos/apuntes/apuntesn82006.pdf

Neisworth, J.T., & Bagnato, S.J. (2005). DEC recommended practices: Assessment. In S. Sandall, M.L. Hemmeter, B.J. Smith, & M.E. McLean (Eds.), DEC recommended practice: A comprehensive guide for practical application in early intervention/early childhood special education (pp. 45–70). Longmont, CO: Sopris West.

Palacios, J., & Lera, M.J. (1992). Escala de Valoración del Entorno Educativo [Evaluation scale for the educational environment. ECERS translation]. Traducción de ECERS. Madrid: Visor.

Perez Juste, R. (2000). La evaluación de programas educativos: Conceptos básicos, planteamientos generales y problemática [The evaluation of educational programs: Fundamental concepts, general frameworks and problematics]. Revista de Investigación Educativa, 18(2), 261–287.

Real Decreto 114 (2004, January 23). Real Decreto por el que se establecen aspectos educativos básicos de educación infantil [Spanish Royal Decree which basic educational contents for pre-kindergarten are established]. BOE 32.

Reynolds, A.J. (2004). Research on early childhood interventions in the confirmatory mode. Children and Youth Services Review, 26, 15–38.

Rivas, S. (2004). Educación temprana en el niño de 0 a 3 años a través de programas [Programbased early education for children between 0 and 3 years old]. Pamplona: EUNSA.

Rivas, S. (April, 2008). Evaluación de la calidad de un programa de educación temprana trilingüe para niños menores de tres años. Conference paper presented at the 15th International Congress of Education and Psychology, Oviedo, Spain.

Rivas, S., & Sobrino, A. (In press). The design and application of a protocol for determining quality of infant–toddler programs in Spain: A case study.

Rivas, S., Sobrino, A., & Peralta, F. (2005). La evaluación como garantía de calidad en educación preescolar [Evaluation as quality standard guarantee in preschool education]. Revista Española de Pedagogía, 63(232), 511–528.

Rous, B., Lobianco, T., Moffett, C.L., & Lund, I. (2005). Building preschool accountability systems: Guidelines resulting from a national study. Journal of Early Intervention, 28, 50–64.

Sarancho, O., & Spodek, B. (2007). Early childhood teachers' preparation and the quality of program outcomes. Early Child Development and Care, 177(1), 71–91.

Schiller, P.B. (2001). Brain research: Its implications for early childhood programs. Child Care Information Exchange, 140, 14–18.

Schwartz, I.S., & Olswang, L.B. (1996). Evaluating child behaviour change in natural settings: Exploring alternative strategies for data collection. Topics in Early Childhood Special Education, 1(16), 82–101.

Shelden, M.L., & Rush, D.D. (2001). The ten myths about providing early intervention services in natural environments. Infants & Young Children, 14(1), 1–13.

Sheridan, S., & Schuster, K. (2001). Evaluation of pedagogical quality in early childhood education: A cross-national perspective. Journal of Research in Childhood Education, 16, 109–125.

Shonkoff, J.P., & Phillips, D.A. (Eds.). (2000). From neurons to neighbourhoods. The science of early childhood development. Washington, DC: National Academy Press.

Sierra, J. (2008). Assessment of bilingual education in the Basque Country. Language, Culture and Curriculum, 21(1), 39–47.

Slavenas, R. (1993). Evaluation practices of model early childhood programs. Early Child Development and Care, 89(1), 31–44.

Stover, D. (2001). Applying brain research in the classroom is not a no-brainer. Education Digest, 66(8), 26–29.

Volpe, J.J. (2001). Neurology of the newborn (4th ed.). Philadelphia, PA: W.B. Saunders. Zalbide, M., & Cenoz, J. (2008). Bilingual education in the Basque Autonomous Community: Achievements and challenges. Language, Culture and Curriculum, 21(1), 5–20.