On the Right Track: Measuring Early Childhood Development Program Quality Internationally

Maria Cristina Limlingan
Tufts University

Two of the main obstacles for drawing educational comparisons consist in determining what are considered 'high quality' initiatives, and finding a common tool that can adapt to differences in both structure and content, as well as to the cultural and demographic characteristics of the population it wishes to serve. This paper focuses on addressing such obstacles by inquiring whether an instrument such as the Early Childhood Environment Rating Scale (ECERS) can be used to compare the early childhood development initiatives in countries like Chile and Bangladesh. Using Pena's (2007) model that considers linguistic, functional, cultural and metric equivalence, we examine the implementation of the ECERS in these two settings, and identify the factors of significance in the instrument's successful adaption to a different context.

Comparisons between early childhood development (ECD) programs offer many potential benefits but are often difficult to execute. Two of the main challenges when drawing comparisons are (1) determining which ECD programs are considered high quality initiatives; and (2) finding a common tool that can adapt to differences in the structure and content of educational systems, as well as the cultural and demographic characteristics of the population it wishes to serve. Ensuring the quality of the program is the primary challenge for the reason that much of the success of an ECD program depends on the quality and approach (Magnuson, Ruhm and Wadofogel, 2007; Myers, 2004). Previous ECD interventions that have produced positive outcomes for children have been designated 'high quality' programs, but little has been written to describe exactly what high quality means. Quality is often relegated to indicators such as teacher-student ratio, teacher education level and teacher experience (in terms of years). These structural factors are often used as indicators of quality because data are obtained easily and studies show that there is a relationship between these factors and childhood outcomes (Peisner-Feinberg & Burchinal, 2001). However, though these structural factors are undeniably useful, they are unable to provide a comprehensive picture of other elements of importance within early childhood settings. Process elements such as teacher-student interaction, learning opportunities and the kinds of activities available are important to look at because these provide a better picture of what is happening, and allow agents to see which areas require increased focus in a child's immediate environment, the very setting which is the most influential and meaningful for the child.

The second challenge is to make comparisons on ECD programs in different cultural contexts. Cross cultural methods enable researchers to test, modify and extend current theories of child development by providing insight into factors in child development that can either be universal or local (Pena, 2007). Applying cross cultural methods allows agents from different regions to gather comparative data in order that it serve as a reference point for the examination of their

respective education systems. Comparisons are also a good way to summarize data that can be easily understood, and it can be used to construct a more persuasive argument during the decision-making process. The availability of information about different early childhood practices has been steadily increasing, however the kind of data collected is usually a mere consolidation of structural elements present in a country's ECD initiative, with little information about how the relative levels of ECD process quality elements compare (Tietze, Bairrao, Leal & Rossbach, 1998).

This paper focuses on addressing these obstacles of determining ECD quality in the international context by inquiring if an instrument such as the Early Childhood Environment Rating Scale (ECERS) can be used to compare the ECD initiatives in countries like Chile and Bangladesh. It will then examine the implementation of the ECERS in these two settings to determine what was considered when the instrument was adapted to a different context.

ECERS and Considerations for Application

Even if educational, socio-economic and cultural differences exist in various regions, certain elements have been recognized as necessary for a child's positive development because of their recurring presence within successful interventions and the literature on the subject. These elements include safe and healthful care, developmentally appropriate stimulation, positive interaction with adults, encouragement of individual emotional growth and the promotion of positive relationships with other children (Tietze et al., 1998). The ECERS covers many of these elements through its seven subscales in the following areas: personal care routines of children, furnishing and display for children, language-reasoning experiences, fine and gross motor activities, creative activities, social development and adult needs. Scores are obtained using a 7-point scale, ranging from inadequate (1) to excellent (7), with each item providing a description of the salient features which need to be observed (Harms, Clifford & Cryer, 1998). Scoring is based on a 2-3 hour classroom observation, and includes a teacher's interview conducted after the observation process. Researchers and practitioners are also encouraged to undergo intensive training to ensure reliability. The ECERS is part of a series of assessments that cover infant-toddler settings (Infant Toddler Environment Rating Scale or ITERS), day care (Family Day Care Environment Rating Scale or FDCERS) and school age children (School Age Children Environment Rating Scale or SACERS). What is unique about the ECERS is its attempt to measure the quality of classroom features enabling the occurrence of pre-identified key processes, whilst simultaneously assessing the processes themselves in the current ECD environment (Villalon et al., 2002). During the creation of the ECERS, Harms, Clifford and Cryer (1980) debated whether to include interpersonal relationships in the scale. The authors relate, however, that they found it impossible to ignore interpersonal behavior and deal adequately with the environment at the same time. The inclusion of items that deal with children's interactions allows teachers and researchers to adopt a more comprehensive means of assessing ECD settings.

As a result of its extensive usage across different countries, changes were made on the ECERS. The adjustments were based on information from a content analysis of the relationship of

ECERS to other global quality instruments, an examination of early childhood program documents, data from studies using the ECERS in preschool and child care settings, and feedback from ECERS users (Harm, Clifford and Cryer, 1998). The result was the Early Childhood Environment Rating Scale Revised Edition (ECERS-R). This version retained basic similarities to the ECERS to ensure continuity, but eliminated some questions to avoid redundancy, provided more detailed descriptions and added items such as health and safety practices, television and computer use and a greater use of interaction questions (Harms, Clifford and Cryer, 1998).

In the United States, results from the ECERS and ECERS –R have been used extensively in research that has examined associations between preschool quality and child development, and dozens of investigations have demonstrated an association between higher scores on ECERS observations and a child's developmental outcomes (Peisner-Feinberg & Burchinal, 1997; Peisner- Feinberg et al., 2001). Results from the ECERS and ECERS – R have also been used to monitor the quality of program(s) and provide guidance for improving quality, and there is ample evidence suggesting that an assortment of program investments, technical assistance, and professional development efforts can be used to improve such scores (Bryant et al., 1994, Howes, Phillips and Whitebook, 1992).

Aside from being used in various capacities in the United States, the ECERS has also been used in other countries where it has been adapted to measure program improvement efforts. Early childhood may exhibit similar elements as being necessary for successful development, but the outcomes in children's development vary and relate in different ways to the many different measures of quality in the ECERS (Sylva et al., 2006). The difficulty in using instruments created in another country is that the standard measures of quality are based on expertise relevant to one region that may not be appropriate for another context. This poses a significant problem if researchers are unable to ensure that the instrument being used retains its relevance and ability to measure in a fair manner. Pena (2007) offers four important features for establishing a study's validity that need to be considered when conducting research across different cultural groups. The first and most commonly known measure, according to Pena (2007), is linguistic equivalence. This refers to translating both instructions and the instrument, and checking to make sure that the words are appropriate for the context in which they are used. Functional equivalence is concerned with ensuring that the instructions and instruments will elicit the same target behavior. Cultural equivalence looks at how respondents will interpret a given direction or test item, and determines if there are possible underlying cultural interpretations that may affect the way an individual responds to the instrument and instruction. Finally, metric equivalence deals with addressing the change in the level of difficulty that might occur (Pena, 2007).

In order to see the effectiveness of the ECERS and the ECERS-R in the international context, we examine the implementation of the instrument in Chile and Bangladesh and look at how the measures were able to address the issues of linguistic, functional, cultural and metric equivalence. Chile and Bangladesh were selected based on their use of the ECERS-R and the

availability of information on how these two countries were able to adapt the instrument to meet their needs.

Chile

Early childhood education has been present in Chile for a long time and increasing its utilization has been one of the priorities of the government (Herrera, Mathiesen, Merino & Recart, 2005). One reason for the uneven coverage may be the fact that even if children from birth to age six can attend ECD programs, attendance is not compulsory (Herrera et al., 2005). Currently, one out of four children from low income families attend ECD programs, compared to one out of two medium or high income children (Herrera et al., 2005). In order to address these issues, the government hoped to first evaluate the existing quality of public preschools by examining how well they met the needs of children and their families.

To carry these goals out, Villalon et al. (2002) conducted a study that compared the different types of preschool (private non subsidized, private subsidized, city council preschool, national program, and those sponsored by non-governmental organizations) that were offered in terms of quality. Using a Spanish version of the ECERS, Villalon et al. (2002) dropped the cultural awareness item which assessed the provision of materials and activities related to diverse context because of its low mean. Ratings were also obtained from 33 experienced pre-school teachers using a questionnaire that asked them to rate quality criterion on a three point scale. In the resulting data, both the *provision of special needs* item and the *naps* item scored very low on the scale. According to Villalon et al., (2002), the reason for the low score may relate to the fact that it is not common practice to have naps within a half day curriculum or to integrate children with disabilities in mainstream early childhood programs. Items such as sand and water, space to be alone, furnishings for relaxation and comfort were rated as relatively important by experts, compared to the remaining 33 items that were rated as very important. Prior to data collection, preschool teachers were trained on the use of the scale until they reached an agreement of 95%. Data collection was done in the middle of the school year in June and July, meaning that classes were settled within a daily routine.

Villalon et al. (2002) found that the average ECERS scores fell into the minimal quality category for the seven subscales with scores that ranged from a low of 3.09 for social development and a high of 4.58 for personal care. Significant differences were found among the two regions (that is, metropolitan and rural areas) in which the study was conducted. Despite differences among six different types of preschool, *personal care routines* and *fine and gross motor skill* areas had the highest average score across the board, while *creativity* and *social development* had the lowest scores.

Bangladesh

Early childhood education programs have grown in popularity with governments in developing countries as a way to prepare the children from high-risk families for school (Myers, 1992). Aboud (2006) confirms that this is the case in Bangladesh, where the early home environment alone is unable to adequately prepare children for school due to the prevalence of

factors such as high malnutrition and low parental education. While it is common for mothers to stay at home, they perceive their role as protecting their child from illness and injury rather than encouraging play and conversation (Aboud, 2006). Children have little exposure to books, media and other educational toys. The purpose of the study by Aboud (2006) was to assess the curriculum typically used in Bangladesh with the objective of developing, through use of different activities, skills related to the process of learning, positive learning attitude and individual learning styles.

Since the study was conducted in a rural setting, a modified version of the ECERS for the South India context called the Tamil Nadu or TECERS (Isley, 2001) was used, together with the ECERS-R. Contextual adaptations were made with the ECERS-R in order to define terms quantitatively. For example, "enough blocks" meant 20 blocks allotted per child, and "some books" meant 10 books. The physical settings used in the TECERS subscale are relevant to a rural setting that has to deal with various indoor and outdoor hazards such as availability of water at toilet and open defecation or urinating. Since the ECD program was a half day program, items concerning meals and naps were excluded. Nine items received the lowest score because there were no televisions, videos, soft toys or cozy areas, and little attempt to protect privacy. Two new subscales were included to address requests arising from parents in the population, and these related to literacy and math activities such as attending to environmental print, emphasizing sounds in words, writing letters and numbers, and counting and matching objects (Aboud, 2006). Twelve research assistants with university degrees were trained for 5 days to conduct testing, and the ECERS-R measures were practiced at nearby schools. Data were collected from October to mid-November, the end of the Bangladeshi school year.

Aboud's (2006) study indicates that the results of ECERS-R ranged from 1.8-3.7 on a 7-point scale, while the converted scores from TECERS ranged from 5.2-5.9 on a 7-point scale. The highest ratings were from the areas of mathematics and literacy, with the lowest scores relating to the areas of activities and program structure, largely due to the scarcity of available fine motor materials.

Comparison between Chile and Bangladesh

In spite of their different contexts, the Chile and Bangladesh studies were able to adapt the ECERS and ECERS-R effectively to suit their needs. The primary goal for linguistic equivalence is to make certain that the words and linguistic meaning used in the instruments and instructions are the same for both versions (Grisay, 2003; Sireci & Berberoglu, 2000). For Chile, the English edition of the ECERS-R was used together with a Spanish translation, Spanish being the language more widely spoken in the country. Simply translating instruments, however, may be insufficient to guard against potential biases and validity threats, and it is important, therefore, to scrutinize the instructions and the choice of words used. In the case of the Chile project, it might have been beneficial to have an expert teacher review the Spanish version of the ECERS before the other teachers were asked to review the tool. In this way, the expert teacher could check the content for differences in word usage between the Spanish used in Spain and that used in a Latin American country such as Chile, differences which may

significantly alter the meaning of the items. Future researchers might also want to consider utilizing the method of back-translation to protect against biases in language. This process consists of a translator who first translates the instrument or instructions from the source language to the target language, then involves another translator who independently translates the target version back to the source language (Pena, 2007). Once both translations are complete, the two versions are compared to identify differences and resolve them. Together with the translation/back-translation, decentering may occur wherein the instrument with translated items may have shifted away from the wording of the source instrument to represent the concept in a manner familiar within the target language. The content resulting from the dual process of back translation and decentering would represent the final version of a tool that is functionally equivalent and linguistically different, yet wherein both versions elicit linguistically similar responses.

Functional equivalence was accounted for by training those who would be using the ECERS to score items consistently. This process was done in Bangladesh, where the researcher and a local Bangladeshi research colleague conducted a five day training course with research assistants who had university degrees. Ensuring the reliability of the research assistants also included having the trainers accompany them during their initial classroom observations and at least one other time during the 6-week data collection. It was noted in the Chile example that the preschool teachers who conducted the classroom observations had previously been trained to use the ECERS, and had reached an agreement of at least 95%.

Metric equivalence is a crucial factor, especially when making comparisons between different instruments. In the Bangladesh study, the researchers decided on using the TECERS because the rural setting of the study made the items on the TECERS more relevant, especially the items relating to the physical environment and personal care and hygiene. However, after using the TECERS in 6 classrooms, the researchers realized that due to the way the TECERS instrument was constructed – with items having a restricted scoring range from 0 – 2 – there was insufficient variability to perform correlations with the ECERS-R. Thus when the TECERS was converted to the ECERS-R rating, the preschool classrooms achieved a 'high quality' designation – that is, exhibiting scores between 5 to 6 on a 7 point scale – which did not seem to correspond to what the researchers had actually observed. Because the TECERS was not metrically equivalent to the ECERS-R the use of the TECERS in the study was discontinued.

Even if items meet the criteria for linguistic, functional and metric equivalence, researchers must also be careful, when considering cultural equivalence, to look and see if items may have salience for different groups due to the distinct cultural and historical ways in which concepts are interpreted by respondents. In the case of the study in Chile, the researchers decided to present a questionnaire with the ECERS items to 33 experienced preschool graduate teachers including those from universities, the ministry of education, directors of preschool centers and preschool practitioners for them to review. The result of this consultation was that certain items such as *nap time*, *children with special needs* and *use of books and videos* were either taken out or modified beforehand, in accordance with the teacher's input, so that the scale would give each

preschool a fair chance with the scores for the ratings. It was interesting to note that the scoring of items as being only *relatively important* amongst the Chilean experts was consistent with the low scores these items obtained in many of the observed Chilean preschool centers, with around 50-78% scoring a 1 in the 7-point scale.

When looking at the results, it is also necessary to understand the cultural context and the outcomes that are important to each specific country. With regard to Chile, similarities found in the ECERS profiles of the different types of preschool assessed demonstrate the influence of shared educational values and orientation in Chile's early childhood education system. Despite significant differences in quality level, some areas were either consistently higher or lower across the entire range of institutions involved, regardless of whether they were private non-subsidized, private subsidized, city council preschool, national program, and non-government organization-sponsored preschool centers. From a practical perspective, what can be taken from the Chile study is the fact that even if there is variation in quality within the type of program, good quality classrooms were identified for each type and this could serve as a model or reference point for raising standards at the classroom level. The challenge then would be decreasing variability within the type of programs and trying to achieve a higher standard.

The Bangladesh project added scales to the ECERS-R such as *literacy*, *math* and *interpersonal interaction* to reflect the importance parents and educators attached to these academic indicators. The fact that the schools scored lower on scales relating to *activities* and *program structure* was largely due to the small variety of challenging materials, and the lack of teacher assistance with individual child progress via hands-on activities and scaffolding. The low results were also consistent with the cultural emphasis on memorization of math phrases and stories rather than on reasoning and vocabulary, a tendency which appears to attenuate the effect of both interest and comprehension amongst the children (Aboud, 2006). Although teachers may be more comfortable teaching by demonstration, these may not be the best methods for enabling children to learn math and language. From these results, certain recommendations were adopted to increase the amount of stimulation children received through materials, activities and instruction (Moore, Akhter, and Aboud, 2008).

In a follow up study by Moore, Akhter & Aboud (2008), curriculum changes in this half day program gave more prominence to language and literacy. Targeted measures, such as daily story reading – with several new stories introduced each week instead of each month – and teaching instructors how to read and talk about stories in an engaging manner (rather than simply requiring memorization) were introduced. The format of the learning was also changed so that it emphasized working in small groups or pairs rather than continuously in small groups. The results of such changes were an increase in ECERS-R scores in *activities* and *program structure* subscales from 3.5 to a score of 4.7 and 6.5 respectively. The increase in scores was also associated with some observable gains in child outcomes (Moore, Akhter and Aboud, 2008). What is even more promising was that the cost to make these changes was estimated at $1.50 per child, per year (Moore, Akhter and Aboud, 2008). In this way, Bangladesh was able to pinpoint and act upon specific areas requiring improvement within their curriculum.

Conclusion

Comparative international research represents an attempt to understand local problems from a global perspective in a world becoming more interconnected and interdependent at all levels (Villalon, Suzuki, Hererra & Mathiesen, 2002). A common instrument is a good way of enriching our knowledge of what is important in different countries. Providing a list of quality measures helps us both to see a more comprehensive picture as well as to make informed decisions about what is important, and what should be modified, added or taken out. For example, mainstreaming special education was not emphasized in Chile, but the government may want to examine this in the future to see if it may be appropriate to their national setting. The rural areas of Bangladesh lacked technology such as television and computers, and this may be an area that they would like to consider developing when planning later on.

Having instruments like the ECERS that measure the quality of early childhood development can serve as a foundation for other countries, especially those that do not have existing measures in place. As demonstrated in Chile and Bangladesh, it is possible to use an instrument such as the ECERS to measure ECD quality internationally. It is important to remember that quality for ECD initiatives must be contextualized in ways that are relevant to the values of a group, and those who plan to use such instruments must be able to take into account linguistic, functional, cultural and metric equivalence. For Chile, the ECERS enabled them to identify areas that were of local and national importance and to compare quality across program types. As a result of this, they may look at ECD programs that are better performing for each type and find ways to improve the quality of their programs. In Bangladesh, they examined a commonly used curriculum in the rural areas. Additional scales were used to adapt to the rural surroundings and the needs of the parents and educators. In the end, they were able to pinpoint the areas that scored the lowest and to make the subsequent decision to work on them to improve child outcomes within a follow-up study.

As the world becomes smaller, it will become more important to find the most effective ways to organize and share information. Cross cultural comparisons using a common instrument, so long as it is composed and utilized in the right way, provides a good method to facilitate discussions which allow us to learn from one another. It is only by doing so that we can hope to avoid making the same mistakes as in the past, and to build a better future for our children.

References

Aboud, F. E. (2006). Evaluation of an early childhood preschool program in rural Bangladesh. *Early Childhood Research Quarterly*, *21*(1), 46-60.

Moore, A.C, Akhter, S., Aboud, F.E (2008). Evaluating an improved quality preschool program in rural Bangladesh. *International Journal of Educational Development*, 28, 118–131.

Bryant, D., Burchinal, M., Lau, L.B and Sparling, J.J (1994) Family and classroom correlates of

Head Start Children's Developmental Outcomes. *Early Childhood Research Quarterly*, 9, 289-309.

Early Childhood Environment Rating Scale (ECERS) Website (2007). Accessed from http://www.fpg.unc.edu/~ecers/ on October 30, 2007

Grisay, A. (2003). Translation procedures in OECD/PISA 2000 international assessment. *Language Testing*, 20, 225–240.

Harms, T., Clifford, R.M.and Cryer D. (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 edition. New York: Teachers College Press.

Herrera, M. O., Mathiesen, M. E., Merino, J. M., & Recart, I. (2005). Learning contexts for young children in Chile: Process quality assessment in preschool centres. *International Journal of Early Years Education*, *13*(1), 13-27.

Howes, C., Phillips, D., & Whitebook, M. (1992). Thresholds of quality: Implications for the social development of children in center-based child care. *Child Development,* 63, 449-460.

Isely, E. (2001). *Tamil Nadu early childhood rating scale*. Chennai, India: M.S. Swaminathan Research Foundation

Myers, R. G., & 20 Consultative Group on Early Childhood Care and Development(1992). *The Twelve who survive : Strengthening programmes of early childhood development in the third world*. High/Scope Press.

Myers, R.G (2004) . *In search for quality programmes of early childhood care and education*. Background paper for Education for All, Global Monitoring Report 2005. Paris: UNESCO. www.unesco.org/education/gmr_download/references_2005.pdf; retrieved December 13, 2007

Magnuson, K., A., Ruhm, C., & Waldfogel, J. (2007). Does prekindergarten improve school preparation and performance? *Economics of Education Review*, *26*(1), 33-51.

Peisner-Feinberg, E., Burchinal, M. (1997). Concurrent relations between child care quality and child outcomes: The study of cost, quality, and outcomes in child care centers. *Merrill-Palmer Quarterly*, 43 , 451-477.

Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., Kagan, S. L., et al. (2001). The relation of preschool child-care quality to children's cognitive and Social developmental trajectories through second grade. *Child development*, *72*(5), 1534-1553.

Pena, E. D. (2007). Lost in translation: Methodological considerations in cross-cultural research. *Child development*, *78*(4), 1255-1264.
Sireci, S. G., & Berberoglu, G. (2000). Using bilingual respondents to evaluate translated-adapted items. *Applied Measurement in Education*, 13, 229 – 248.

Sylva, K., Siraj-Blatchford, I., Taggart, B., Sammons, P., Melhuish, E., Elliot, K., & Totsika, V., (2006). Capturing quality in early childhood through environmental rating scales. Early *Childhood Research Quarterly*, *21*, 76-92.

Tietze, W., Bairrao, J., Leal, T. B., & Rossbach, H. (1998). Assessing quality characteristics of centerbased early childhood environments in Germany and Portugal: A cross-national study. *European Journal of Psychology of Education*, *13*(2), 283-298.

Villalon, M., Suzuki, E., Herrera, M. O., & Mathiesen, M. E. (2002). Quality of Chilean earlychildhood education from an international perspective. *International Journal of Early Years Education*, *10*(1), 49-59.