Preschool Experience in 10 Countries: Cognitive and Language Performance at Age 7

Jeanne E. Montie, Zongping Xiang, & Lawrence J. Schweinhart

High/Scope Educational Research Foundation, 600 N. River St.,

Ypsilanti, MI 48198, USA

Montie, J. E., Xiang, Z., & Schweinhart, L. J. (2006). Preschool experience in 10 countries: Cognitive and language performance at age 7. *Early Childhood Research Quarterly, 21,* 313–331.

Contact information: jmontie@highscope.org or lschweinhart@highscope.org

Abstract

The IEA Preprimary Project is a longitudinal, cross-national study of preprimary care and education designed to identify how process and structural characteristics of the settings children attended at age 4 are related to their age-7 cognitive and language performance. Investigators collaborated to develop common instruments to measure family background, teachers' characteristics, setting structural characteristics, experiences of children in settings, and children's developmental status. Data from 10 countries are included in the analysis; in most countries, the sample of settings is representative of preprimary settings in that country. For the analysis, a 3-level hierarchical linear model was employed that allowed decomposition of variation of child outcomes into 3 parts—variation among children within settings, among settings within countries, and among countries. Four findings are consistent across all of the countries included. Age-7 language improves as teachers' number of years of full-time schooling increases and the predominant type of activity teachers propose in settings is free choice rather than personal/social. Age-7 cognitive performance improves as children spend less time in whole group activities and the variety of equipment and materials available increases. There were also a number of findings that varied across countries depending on particular country characteristics. The findings support child-initiated activities and small group activities and are consistent with developmentally appropriate practices promoting active learning.

Preschool Experiences in 10 Countries: Cognitive and Language Performance at Age 7

The wide range of environments throughout the world in which young children grow and learn creates challenging questions for everyone concerned with providing high-quality programs for preprimary children. What are the essential program elements that promote optimum child development? How can they be measured or assessed? How are these elements delivered in various types of settings? The International Association for Evaluation of Educational Achievement (IEA) sponsored the Preprimary Project to provide information to help early childhood educators and policy planners in various countries answer these questions. The goal of this cross-national project was to identify and document the process and structural characteristics of care and education settings for 4-year-olds and relate these characteristics to the children's cognitive and language performance three years later.

Sampling in most early childhood research studies is limited to one, usually western industrialized, country. Only a small number of existing studies are based on nationally representative samples. The Preprimary Project sample includes diverse countries and, in most countries, the sample of settings is representative of the country's total population of settings. The combination of the international sample and extensive data collection creates a rare opportunity to explore questions that have not been previously addressed in the early childhood research literature. For example, do similar teaching practices lead to common outcomes in diverse countries? Do structural characteristics, such as group size, have a similar relationship to child outcomes in countries with different child-rearing philosophies?

Background

Research has demonstrated that high quality early childhood education and care can have lasting, positive outcomes both for children from disadvantaged backgrounds and children from nondisadvantaged backgrounds (Campbell, Ramey, Pungello, Sparling, & Miller-Johnson, 2002; Peisner-Feinberg, et al., 2001; Reynolds, Temple, Robertson, & Mann, 2001; Schweinhart et al., 2005); however, there is a lack of cross-national research that identifies fundamental principles of high quality in programs in diverse countries. The Preprimary Project was conceived in the mid-1980s to provide data to address this question. The IEA invited member countries to participate in one or more of 3 phases. In Phase 1, a household survey was administered to families to identify the major early childhood care and education settings used by families with 4-year-old children (Olmsted & Weikart, 1994). In Phase 2, an observational study documented the structural and process characteristics of selected setting types identified in Phase 1 and assessed children's cognitive and language skills at age 4 (Olmsted & Montie, 2001; Weikart, 1999; Weikart, Olmsted, & Montie, 2003). In Phase 3, the children observed in Phase 2 were followed to age 7, when language and cognitive assessments were administered. In all, 17 countries participated in one or more phases of the project. The results reported in this article are based on the 10 countries that have complete data sets for Phases 2 and 3. The participating countries, listed in Table 1, were from North America, Europe, and Asia and include both developed and developing nations.

A large body of research carried out in western countries seeks to understand the contributions of specific structural and process characteristics to the quality of early childhood settings by examining the relationships of various characteristics of settings to measures of global quality and children's outcomes. In a review of the longitudinal associations between structural and caregiver characteristics and children's academic, cognitive, behavioral and social development, Vandell and Wolfe (2000) concluded that "smaller group sizes, lower child-caregiver ratios, and more caregiver training and education appear to have positive effects on …important developmental outcomes" (p. 22).

Process characteristics are more difficult to measure than structural characteristics; two approaches can be used—direct observation of teachers' and children's behaviors in the classroom or global rating scales such as the Early Childhood Environment Rating Scale (ECERS, Harms, Clifford, & Cryer, 1998). One cross-national study involving 4 countries – Germany, Portugal, Spain, and the U.S. – sought to predict process quality, measured by the ECERS and the Caregiver Interaction Scale (Arnett,1989), from structural characteristics such as adult/child ratios, teacher qualifications and wages, and regional characteristics (Cryer, Tietze, Burchinal, Leal, & Palacios, 1999). The study found that although many of the same structural characteristics were related to process quality, across countries, no one structural characteristic strongly predicted process quality. The authors concluded that many structures work together to create quality and improving process quality may require the direct regulation of program processes (e.g., actual classroom practice).

The Observational Study of Early Childhood Programs (Layzer, Goodson, & Moss, 1993) demonstrated the potential of direct observation of process characteristics in preprimary classrooms to inform teaching practice and classroom organization in a way that measures of global quality cannot. The investigators used both rating scales, to measure global quality, and direct time-sampling observations, which yielded detailed descriptions of classroom activities and groupings, teacher behaviors and interactions, and child activities and interactions. They concluded that the direct observations were better predictors than measures of global quality of the child behaviors (e.g., task engagement, use of higher level social strategies) they defined as child outcomes.

The Preprimary Project differed from previous research in that the international research team made few assumptions as to what constituted high-quality programs. When selecting the setting characteristics to be examined, project researchers defined the range of experiences that children have in settings in different countries and attempted to include all major variables representing setting structure and process that were part of children's experiences. Research available when the Preprimary Project was conceived -- the U.S. National Day Care Study (Ruopp, Travers, Glantz, & Coelen, 1979), the Oxford Preschool Research Project (Sylva, Roy, & Painter, 1980), and Katz's research on teacher behaviors (Katz, 1968) -- suggested various aspects of setting structure and process to be examined. These included group size; staff/child ratio; teachers' education, experience, and organization of activities; children's interactions with adults and other children; and children's activities.

The Preprimary Project used variables derived from direct observation of children's and teachers' behaviors and activities in preschool, as well as setting structural characteristics, to

predict children's language and cognitive performance at age 7. The aims of the analysis were twofold: (a) To examine whether children's experience in settings at age 4 was related to their language and cognitive performance at age 7 and (b) to examine whether relationships between children's experience in settings and their performance at age 7 varied across participating countries and to explore reasons for varying effects.

Method

Sampling Procedures

The target population for Phase 2 was children in selected early childhood settings who were between the ages of 4 years 3 months and 4 years 9 months. The population of settings to be sampled in Phase 2 was guided by the results of the Phase 1 household survey study. Researchers in each of the participating countries worked with the international sampling referee[1] to select one or more types of early childhood settings (e.g., public preschools, private child care centers) that were used by large numbers of families or were important for other public policy reasons. Each country's sample was representative of settings in the country or major areas of the country. Whenever possible, settings were selected from a national list of all such settings to obtain a nationally representative sample. When no lists were available, area sampling procedures were followed.

Table 1 provides information about the Phase 2 and 3 achieved samples of settings and children by type of setting in each participating country. The original sampling plan was designed to achieve sample sizes sufficient for individual country multivariate analyses at the type of setting level (assuming 20% attrition). To achieve sufficient variation in settings, the goal was to select a minimum sample of 24 settings of each type and 96 children, thus, a maximum of 4 children were randomly selected from each classroom Detailed information about sampling procedures used in each country is available in Weikart et al. (2003).

The median retention rate from Phase 2 to Phase 3 across types of settings within countries was 89%, with a range of 49% to 99% (see Table 1). In some countries (e.g., Italy), funding problems limited Phase 3 data collection efforts, while in others, although a variety of efforts were used to locate children, some were lost to follow up. Phase 2 data were collected primarily during 1992. The data were collected at different times during the school year in different countries for practical reasons, such as varying recruitment schedules and differences in school schedules. Phase 3 data were collected primarily in the last quarter of the 1994-95 school year.

Instrumentation and Data Collection

Phase 2 instruments consisted of 3 observation systems and 3 questionnaires (see Table 2), as well as age-4 developmental status measures. Cross-national teams defined specific areas of interest and the variables associated with them, reviewed existing instruments, and collaboratively developed the data collection instruments used in all the countries. The use of rating scales for process measures was rejected for two reasons: (a) rating scales produced in developed countries were deemed inappropriate for use in developing countries, and (b) researchers wanted to focus on specific teacher and child behaviors and activities. All of the instruments were developed with substantial input from researchers in all of the countries

involved; two rounds of pilot testing took place in each country, with revision in between as necessary.

Observation Systems

Observations were used to measure process characteristics, that is, to collect information about what children and adults were doing and how children's time was organized by adults. The Child Activities (CA) system was used to document the activities and interactions of the target child in a given setting. The observations were coded into a predetermined category system that included 12 major categories (see Table 2). Most categories were further broken down into subcategories; for example, expressive activities could be further coded as dramatic play, arts and crafts, or music. Children's interactions with adults and other children were also recorded. The Adult Behavior (AB) observation system was used to document the behaviors of the primary adult in a setting. Major categories (see Table 2) were broken down into subcategories; for example, teaching could be further coded into 11 subcategories (e.g., giving demonstrations, eliciting information, offering choices, etc.), which were then classified as adult-centered or child-centered teaching. Each adult and target child was observed for two 10-minute periods in the morning on two nonconsecutive days, yielding a total of 40 minutes of observation for each. The Management of Time (MOT) system documented how the primary adult in a setting organized children's time as a running record for 1½ to 3½ hours during 2 nonconsecutive mornings. The activities were then coded into a predetermined category system that included 12 major categories that were similar to those in the CA system with the addition of two categories--free choice and mixed activities (different activities proposed for small groups of children). Group structure, whether the activity was proposed for the whole group, a small group, one or two children together, or one child alone, was also coded.

The observation system findings were calculated as percentages of total time for each category, and means were calculated by type of setting. A description of each of the observation instruments and the coding categories and descriptive findings from all 3 observation systems by country and type of setting can be found in Weikart et al. (2003).

Questionnaires

Three questionnaires were used to collect information about the structural features of the early childhood settings, the target child's family background, and the expectations of teachers and parents about what is important for preprimary children to learn (see Table 2). A detailed description of the Expectation Questionnaire and Phase 2 findings is presented in Weikart (1999). The Family Background Interview collected information about the family characteristics of the target child. The Provider Survey was administered to setting administrators and teachers to gather information about structural characteristics of the setting. The findings from this survey are presented and discussed in Olmsted and Montie (2001).

Developmental Status Measures

Each of the 4 measures--language at ages 4 and 7 and cognitive at ages 4 and 7--was developed by a cross-national team using a combination of items from existing tests. First, items were translated from English into each country's dominant language, back-translated to check for translation accuracy, and then pilot-tested in each country. The national research teams reviewed the pilot-test results to determine that the order of difficulty was similar for all countries, the pass rates for individual items were in an appropriate range and did not differ

greatly across countries, and the scoring systems accurately reflected the children's proficiencies. If translations and illustrations used in the items were problematic in a particular country, modifications that did not alter basic content were made as necessary to make the items appropriate for use in that country. The revised tests were pilot-tested and after a review of the pilot-test data, minor changes were made as necessary to insure clarity and uniformity.

The Phase 2 age-4 language measure[2] tested receptive and expressive language and required the child to perform a variety of tasks such as telling stories, answering questions, matching pictures to words or phrases, and repeating statements. The Phase 2 cognitive measure[3] assessed the child's knowledge and skills in three content areas--spatial relations, quantity, and time. Children were required to demonstrate understanding of a wide variety of concepts by performing an action, pointing to a picture, or responding verbally. The Phase 2 developmental status measures were administered at different times during the school year in different countries; thus, they were not used to measure outcomes of preprimary experience, but instead established a baseline that was used to adjust age-7 outcomes for prior ability levels.

The Phase 3 age-7 language measure[4] asked the child to perform a variety of tasks such as telling stories, answering questions, matching pictures to words or phrases, ordering sentences, and repeating statements. The Phase 3 cognitive measure[5] assessed the child's knowledge and skills in 5 content areas: spatial relations, quantity, time, memory, and problem solving.

After the assessment data were collected, independent testing experts examined the results to determine their suitability for use as developmental assessment tools in multiple countries (detailed results are presented in Wolfe & Manalo, 2005). After determining that the cognitive and language instruments were each unidimensional and that each tested a different construct, the items were scaled using latent trait modeling, specifically the three-parameter logistic model (Hambelton & Swaminathan, 1985). This model describes the influence of 4 parameters--the child's ability, the item's difficulty, the item's discrimination, and the opportunity to guess the correct answer--on the probability that a child will answer a specific item correctly. Reliability estimates for the 4 tests using the reliability of separation index (analogous to Cronbach's coefficient α), ranged from .81 to .94. The researchers also evaluated the functional equivalence of the multiple language versions of the instruments using differential item functioning statistics to depict the degree to which individual items within a test function differently as construct measures for different groups of children (i.e., those who take different language versions of the same instrument). Results showed that an unacceptably large number of items on each of the tests were not functionally equivalent; in spite of careful translation, many of the items did not function in the same manner in different countries. The researchers concluded that the Phase 2 and Phase 3 language and cognitive measures were reliable, valid, and suitable for use as developmental assessment tools in diverse countries. However, the functional inequivalence of items across countries precluded the possibility of comparing the test results across countries.

Training Procedures and Reliability

A common set of procedures was used to train data collectors in all participating countries. The representatives who supervised the training and data collection in each country

were first trained by researchers from the study's International Coordinating Center (ICC) and then trained data collectors in their own country. Data collectors were persons with experience in early childhood, such as teachers or graduate students in the field. Each one had to reach or exceed 80% agreement with a standard before being certified to collect data on each of the instruments described above. Data coding and entry took place in the country where data were collected or, if that capability did not exist, data were sent to the ICC for coding and entry. All data analysis took place at the ICC.

Analysis

Statistical Approach

Based on the structure of the Preprimary Project data, with individual children nested within settings and settings nested within countries, hierarchical linear modeling (HLM) was employed for the analysis (Bryk & Raudenbush, 1992). The advantage of HLM lies in its accurate estimation of effects for different levels of variables, which was especially important for this project because effects of settings and countries were often confounded with one another, for example, due to the functional inequivalence across countries in the age-7 outcome measures. The use of an HLM 3-level approach permits the decomposition of the variation in child outcomes into 3 parts--variation among children within settings, among settings within countries, and among countries. As a result, relationships between setting variables and children's age-7 test scores, the primary focus of the study, are free from substantial influence of country as well as child-level effects. Another advantage of HLM for this study is its capacity to explore reasons for heterogeneous relationships across countries by modeling regression coefficients as a function of country-level variables, for example, why amount of adult-child interaction was positively related to child outcomes in some countries, but not or negatively related in others. Examination of the country-level variables that predicted this heterogeneous relationship can shed light on the nature and effects of adult-child interaction in different cultural contexts.

Variables

The two outcome variables for the study were children's age-7 language and cognitive performance scores.

There were 6 predictors included at level 1, the child level. Their purpose was to adjust for differences among children, while estimating the setting effects that lasted through the early elementary years. The predictors included child demographic characteristics (age, age spread [months between age-4 and age-7 tests], gender, and number of siblings); socioeconomic status (parents' averaged years of education); and prior developmental status (age-4 test scores). The level-1 predictors were selected because of the design of the study (age-4 scores), evidence of their association with child development in either this study (age, age spread, number of siblings, and parents' education) or in the research literature (gender).

Level-2 predictors, variables of major interest, represented children's preschool experience at age 4 (see Table 3). The level-2 setting variables represent both process (activities and interactions) and context (structural characteristics). The process predictors (the first 9 variables listed in Table 3) were generated from the 3 observation systems, including 2 MOT variables (predominant activity, represented by 4 dummy variables indicating which of 4 major

activities was proposed most frequently in the setting, and predominant group structure, represented by amount of whole group activity); 4 CA variables (amount of adult-child interaction, child-child interaction, children's group response, and no active engagement); and 3 AB variables (adults' characteristic behavior represented by 3 dummy variables indicating which of 3 behaviors was characteristic of the adult, amount of adult listening to children, and amount of adult-centered teaching). The structural characteristics of settings were represented by 4 variables--group size, number of years of teacher's education, materials available in the setting, and mean years of parents' education, used as an indicator of the general resources of a setting and generated from child-level data averaged for each setting.

No variables were originally designed and collected at the country level. Instead, level 3 variables (see Table 3) were generated by aggregating data from each country's sample to serve as indicators of the major characteristics of a country's preprimary care and education--practice, resources, and values. Country practice was represented by the same setting process predictors used at level 2 and country resources were represented by the same predictors of setting structural features used at level 2. Variables representing a country's values were derived from averaged teachers' rankings of what they believed was important for preschool children to learn in 3 areas--preacademic skills, language skills, and social skills with peers. These generated country-level predictors were not intended to explain the variation in test scores across countries; it was impossible to disentangle the effects of any country variables on the test scores from effects due to functional inequivalence of the outcome measures. However, relationships found *within* a country were assumed not to be contaminated by functional inquivalence, and, therefore, were legitimate for further exploration.

Sample

Because data for setting structural features were not available for 3 countries (Greece, Indonesia, and Spain), two separate analyses were conducted for each outcome variable: a 10-country analysis that omitted the 3 structural feature predictors (group size, teachers' years of education and materials available for children), and a 7-country analysis that included the 3 features. Because of the smaller number of countries included, country-level predictors were not included in the 7-country analysis, which was performed solely to explore findings for the 3 structural features of settings.

Table 4 lists the countries included in the two analyses and sample size at both the child- and setting-level, overall and by country. The mean test age per country ranged from 4.5 to 4.8 for Phase 2 and 7.5 to 8.2 for Phase 3. The percentage of males per country ranged from 43% to 54%. Preliminary analyses indicated that in all countries except Ireland, the variance across types of settings was smaller than the variance within setting types. To insure approximately equal weighting for each country, only the odd numbered settings in Poland were included. The final setting sample size ranged from 47 to 68 for all but 2 countries. In Ireland (setting $n = 97$), large differences were found between the 2 major types of settings included in the sample--national schools and private preschools. In the U.S. (setting $n = 78$), 3 types of group settings were included (Head Start, public preschools, other group settings), compared to 1 or 2 types of settings for other countries. Overall, the total number of settings included was 632 for the 10-country analysis, and 426 for the 7-country analysis.

The average number of children per setting was 3 for 8 out of 10 countries, but less than 2 for Finland and Greece. Overall, 47% of all the settings had 4 children, 25% had 3, 15% had 2 and 13% had 1. The total number of children was 1,897 for the 10-country analysis, and 1,300 for the 7-country analysis.

Model Specification

Level 1. We modeled children's age-7 score as a function of 6 level-1 predictors:

$$Y_{ijk} = \pi_{0jk} + \pi_{1jk}\text{Age 4 Score}_{ijk} + \pi_{2jk}\text{ Age}_{ijk} + \pi_{3jk}\text{Age Spread}_{ijk} + \pi_{4jk}\text{Gender}_{ijk} + \pi_{5jk}\text{Siblings}_{ijk} + \pi_{6jk}\text{Parents' Education}_{ijk} + e_{ijk}. \quad \text{(Level 1)}$$

Where Y_{ijk} is age-7 test score of child i in setting j in country k; π_{0jk} is setting jk's mean score (intercept) adjusted for children's differences in the 6 predictors based on the entire sample; $\pi_{1jk}, \pi_{2jk}, \ldots \pi_{6jk}$ are regression coefficients (slopes) of the corresponding predictors in setting jk; and e_{ijk} is the child-specific random effect.

Level 2. For each of the 6 setting-specific slopes, given the small number of children the slope was based on, we chose to use the country's mean regression coefficient as its estimate for each setting within country. The specific equation for the pth slope in setting jk was therefore:

$$\pi_{pjk} = \beta_{p0k}. \quad \text{(Level 2a)}$$

Where π_{pjk} is regression coefficient p in setting jk; β_{p0k} is the mean regression coefficient of p for country k.

For the setting-specific intercept (setting's mean score), we modeled its variation across settings as a function of all the setting-level predictors listed in Table 2 as the initial step (without the 3 setting structural features in the 10-country analysis). Setting-level predictors that were not related to the setting's mean score and whose effects did not vary across countries were dropped from the initial equation, except for a setting's mean parent education, which was always included as a control variable for a setting's general resources. The final equation for predicting variation in settings' mean scores only included setting-level predictors for which the regression coefficient was either statistically significant or varied across countries, plus the setting's mean parent education and a random effect:

$$\pi_{0jk} = \beta_{00k} + \beta_{0qk}\text{'s Setting-level predictors}_{jk} + r_{0jk} \ (q = 1, 2,\ldots q). \quad \text{(Level 2b)}$$

Where π_{0jk} is the mean score for setting jk; β_{00k} is country k's mean score; β_{0qk}s are regression coefficients of the corresponding setting-level predictors qs (q = 1, 2,…q) in modeling the setting's mean score in country k; and $r0_{jk}$ is the setting-specific random effect.

Level 3. For each of the country-specific slopes (6 for level-1 predictors, β_{p0k}s; and the remaining ones for level-2 predictors, β_{0qk}s), we modeled its variation across countries as random first, and then constrained it to be zero if no cross-country variation was detected. The equation for the *pqth* homogeneous slope in the *kth* country was therefore:

$$\beta_{pqk} = \gamma_{pq0}. \qquad \text{(Level 3a)}$$

Where β_{pqk} is the country *k*'s regression coefficient *pq* found to be homogeneous across countries; and γ_{pq0} is the mean regression coefficient of *pq* for all the participating countries. For country-specific slopes that were found to vary across countries, we then modeled each of them as a function of the country-level predictors listed in Table 2. Given the limited number of countries in the study, level-3 predictors were included in the model one at a time. Each of the final equations for modeling heterogeneous slopes had only one or two country-level predictors, plus a random effect. The specific equation for the *pqth* heterogeneous slope in the *kth* country was therefore:

$$\beta_{pqk} = \gamma_{pq0} + \gamma_{pqm}(\text{s})\ \text{Country-level predictor(s)}_k + U_{pqk}. \quad (m = 1, 2) \qquad \text{(Level 3b)}$$

where β_{pqk} is country *k*'s regression coefficient *pq* that varied across countries; γ_{pq0} is the mean regression coefficient of *pq* for all the participating countries; γ_{pqm}(s) is the regression coefficient(s) of the country-level predictor *m*(s) (*m* = 1, 2) in predicting the heterogeneous country-specific coefficient *pq*; and U_{pqk} is the country-level random effect on *pq*. In fact, no random effect (U_{pqk}) was statistically significant after country-level predictors were included for each of the level-3b equations. We kept it as random to avoid overestimating the power of estimates for the country-level predictors. (Note no country-level predictors were included in equations level-3b in the 7-country analysis.)

For the country-specific intercept (country's mean score), we chose to model its variation across countries as a function of country dummy variables:

$$\begin{aligned}\beta_{00k} = {} & \gamma_{000} + \gamma_{001}\text{Country1}_k + \gamma_{002}\text{Country2}_k + \gamma_{003}\text{Country3}_k + \gamma_{004}\text{Country4}_k \\ & + \gamma_{005}\text{Country5}_k + \gamma_{006}\text{Country6}_k + \gamma_{007}\text{Country7}_k + \gamma_{008}\text{Country8}_k \\ & + \gamma_{009}\text{Country9}_k. \end{aligned} \qquad \text{(Level 3c)}$$

Where β_{00k} is country *k*'s mean score; γ_{000} is the mean score for all the participating countries, and γ_{001} to γ_{009} are the corresponding country effects on country's mean score compared to the referenced country. There were two reasons for explicitly controlling for all the unobserved differences in countries' mean scores: one was to insure that no country effects, including that of functional inequivalence of the tests, were involved in estimation of setting effects at level 2; the other was to achieve less biased estimates for country-level predictors at level 3. As a further test, we modeled variation in country-specific intercepts and country-specific slopes using several different specifications, including constraining the error term to be zero in equation level

3b as a check of non-distorted estimates for the included country-level predictors. The different specifications of the country-level equations did not substantially affect estimates for the included country-level predictors, with the results using the current intercept equation being the closest to the non-distorted estimates for the country-level predictors in the slope equations.

Results

Table 5 presents descriptive statistics for the outcome and predictor variables at each level. Cognitive and language scores were standardized with a mean of 0 and standard deviation of 1. The first 4 variables listed for level 2 -- MOT free choice, MOT physical/expressive, MOT preacademic, and MOT personal/social -- are dummy variables, which together represent the predominant activity in the setting. Level-2 variables AB teaching, AB participation, and AB non-involvement are dummy variables, which together represent the predominant adult behavior in the setting. Level-3 variables are averaged across the 10 participating countries. Preliminary analyses showed that intercorrelations among predictors at levels 1 and 2 were generally weak (r = .10 - .20) with two exceptions. At level 1, age and age-spread were correlated at .68 and at level 2, adult-child and child-child interaction were correlated at .48. At the country level, intercorrelations were moderate to strong, ranging from .30 to .80.

Tables 6 and 7 present results of the HLM analysis for children's age-7 language and cognitive performance, respectively. Predictors whose fixed effects are not presented in the outcome tables, but are listed in Table 2, were dropped from the final model due to lack of evidence of their relationship to children's age-7 scores. All the estimates, both fixed and random effects, are based on the 10-country analyses except the fixed effects for two setting structural features, teachers' education in Table 6, and materials in Table 7, which are based on the 7-country analyses. Fixed effects are presented in two groups, effects for adjustment and effects of interest, with the latter as the focus of the following discussion. The primary focus of the analysis is the relationship between level-2 predictors, including both setting process and structural characteristics, and children's age-7 scores. Level-1 child and family predictors are used for adjustment only. Level-3 country dummy variables are used to adjust for country-level effects; therefore, their estimates are not presented because they may have been distorted due to the functional inequivalence of the tests. For the random effects, variance components for each of the three levels in the initial unconditional model are presented first, followed by other parameters. Percentage of unexplained variance was obtained by subtracting the unexplained variance (amount left unaccounted for after all the predictors were included in the final model) from the unconditional variance (amount before inclusion of the corresponding level of predictors), and dividing by the unconditional variance.

S*etting Process and Structural Characteristics and Children's Age-7 Language Scores*

The initial decomposition of variance in children's age-7 language scores showed that 22% of the total variance was from countries, 23% was from settings, and 55% was from individual children. Several significant relationships were found among the predictors used for adjustment and are listed in Table 6. As expected, children's age-4 language scores were positively and strongly related to their language scores at age 7, on average. The magnitude of this effect was related to a country-level predictor (AB participation_3); thus, the strength of the relationship between children's age-4 and age-7 language scores depended on a country's

percentage of settings where the characteristic adult behavior was participation in children's activities. This variable may covary with a country's educational practice or beliefs about children, which were not measured directly in this study.

Overall, the six level-1 child and family predictors explained 18.3% of the child-level variance in their age-7 language scores, and the level-2 setting predictors explained 17.1% of the variance available for setting-level predictors, as shown in the lower half of Table 6 under random effects. The significant amount of variance left unexplained (.142, $p < .0001$) indicates that there are still important setting characteristics and other variables that were not included in the analysis, for example, those related to the primary schools children attended prior to age 7. Although specified as not applicable for technical reasons, the entire variance in countries' mean scores was explained by the country dummy variables. With adjustment for both country and child effects on children's age-7 language scores, the following sustained relationships were found between level-2 setting predictors and children's language performance at age 7.

Predominant Activity

Children in settings in which free choice activities (teachers allow children to choose their own activities) predominated achieved a significantly higher average language score at age 7 (.18 point of the standardized score) than their counterparts in settings in which personal/social activities (personal care and group social activities) predominated and a nearly significantly ($p = .06$) higher score than their counterparts in settings in which preacademic activities predominated. No significant difference was found between physical/expressive and free choice activities. These findings did not vary across countries.

Adult-Child Interaction

On average, the amount of adult-child interaction in settings was not related to children's language performance at age 7; however, this effect was found to vary across countries. The magnitude and direction of the effect depended on two country-level predictors – a country's average amount of group response and adult-centered teaching. In countries with a high rate of group response (1 *SD* above the grand *M*), the effect of adult-child interaction decreased by .11 point. This resulted in a decrease in the standardized score of 0.15 point (-.04 + 1 x -.11) for every standard deviation increase in adult-child interaction, while controlling for the country's average amount of adult-centered teaching. If a country's average rate of group response was 1 standard deviation below the grand mean, the adult-child interaction effect increased by 0.11 point, resulting in a total increase in children's language scores of .07 point [-.04 + (-1) x -.11] for every standard deviation increase in adult-child interaction, while controlling for average amount of adult-centered teaching. The interpretation for the effect of a country's average amount of adult-centered teaching is similar. With a country's average amount of group response and adult-centered teaching both taken into account, the effect of adult-child interaction in settings ranges between -.26 (-.04 + 1 x .11 + 1 x -.11) and .18 [-.04 + (-1) x -.11 + (-1) x -.11]. Thus, the amount of adult-child interaction was positively related to children's age-7 language scores in countries where group response and/or adult-centered teaching were relatively infrequent, and negatively related to their language scores in countries where group response and adult-centered teaching were frequent.

Child-Child Interaction

The amount of child-child interaction in settings, though not related to age-7 language scores on average, was found to vary significantly across countries depending on a country's amount of whole group activities and teacher beliefs about the importance of child language development. The effect of whole group activities was negative (-.17, $p < .01$), while the effect of teacher beliefs about the importance of language was positive. With effects of both country-level predictors taken into account, the effects of child-child interaction ranged between -.36 and .26. Thus, the amount of child-child interaction was positively related to children's age-7 language scores in countries where whole group activities were relatively infrequent and/or teacher beliefs about the importance of language development were relatively high, but negatively related to their age-7 language scores in countries where whole group activities were frequent and/or teacher beliefs about the importance of language development were relatively low.

Teachers' Education

Based on the 7-country analysis, the number of teachers' years of full-time education was positively related to children's age-7 language scores. Every standard deviation increase in teachers' full time education resulted in 0.07 point rise in children's age-7 language scores. The two other structural variables included in the analysis, group size and amount of materials, were not significantly related to children's age-7 language scores.

Setting Process and Structural Characteristics and Age-7 Cognitive Scores

The initial decomposition of the variance in children's age-7 cognitive scores showed that 66% was from individual children, 26% was from settings, and only 8% was from countries. Thus, less variance in children's cognitive scores than in their language scores came from countries, but considerably more variation came from individual children. This implies that children's cognitive development was less affected by national, social, and cultural factors, and more by child and family characteristics. Compared to the results for children's age-7 language scores, greater effects were found for the two age-related predictors, which implies that children's growth plays a greater role in cognitive than language development. Their cognitive scores at age 4 were strongly and positively related to their age-7 scores, and the magnitude of this effect was related to 2 country-level predictors: the average amount of time adults spent teaching and the percentage of proposed preacademic activities. In countries with little adult teaching and/or frequent preacademic activities, children's age-4 cognitive scores were more strongly correlated to their age-7 cognitive scores. In countries with lots of adult teaching and/or fewer proposed preacademic activities, their age-4 scores were less strongly correlated with their age-7 scores. These country-level predictors may covary with characteristics of a country's educational system and thus reduce the differentiating effects of children's background on their later development; however, a discussion of these effects is beyond the scope of this article.

As shown in Table 7, under random effects, 14.8% of the variance at the child level was explained by the child-level predictors, and 8.8% of the variance available for setting-level predictors was explained by setting-level predictors. The significant amount of variance left unexplained indicates that important setting characteristics and other variables (e.g., primary schools attended by children) were not included in the analysis.

After the variables for adjustment were included, the following relationships were found between children's age-4 experience in settings and their cognitive scores at age 7.

Whole Group Activities

The amount of whole group activities adults proposed in settings was negatively related to children's cognitive performance at age 7. The coefficient for whole group activities in Table 7 (-0.10) indicates that as the amount of whole group activities increased by 1 standard deviation, children's age-7 cognitive scores decreased by .10 point of the standardized score. Thus, children in settings with a low frequency of proposed whole group activities scored higher, on average, than their counterparts in settings with a high frequency of whole group activities. No variation in this effect was found across the 10 countries.

Adult-Child Interaction

Similar to the findings for the language analysis, on average, the amount of adult-child interaction was not related to cognitive performance at age 7; however, the effect varied across countries depending on a country's percentage of settings in which free choice activities predominated (free choice_3 = .18). In countries where the percentage of settings in which free choice predominated was 1 standard deviation above the grand mean, every standard deviation increase in adult-child interaction in settings resulted in a .18 increase in children's age-7 cognitive scores. In contrast, in countries where the percentage of settings in which free choice activities predominated was 1 standard deviation below the grand mean, children's age-7 cognitive scores dropped by 0.18 point for every standard deviation increase in adult-child interaction, thus, the effect of adult-child interaction in settings ranges from -.18 to .18.

Variety of Materials

Based on the 7-country analysis, the variety of materials available for children's use in settings was positively related to children's cognitive performance at age 7. Every standard deviation increase in materials resulted in a 0.09 point increase in children's age-7 cognitive scores. Teachers' education and group size were not related to children's age-7 cognitive scores.

Discussion

Findings that Do Not Vary Across Countries

Four findings apply to all the countries included in the study.

1. Children who were in preprimary settings in which free choice activities predominated had significantly better language performance at age 7 than those in settings in which personal/social activities predominated.
2. As level of teacher education increased, children's age-7 language performance improved.
3. The less time children spent in whole group activities, the better was their age-7 cognitive performance.
4. As the number and variety of materials in settings increased, children's age-7 cognitive performance improved.

The positive relationship between the number of years of teachers' education and children's language scores makes intuitive sense and is supported by research in both home and

preschool settings. Research shows that parents and teachers with higher levels of education use more words and more complex language when communicating with children (Dickinson, 2001b; Hart & Risley, 1995). Researchers in the United Kingdom found that more highly trained teachers were the most effective in their interactions with children, using the most sustained shared-thinking interactions (Siraj-Blatchford, Sylva, Muttock, Gilden, & Bell, 2002). The positive relationship between the availability of a wide range of materials and children's cognitive scores also makes intuitive sense. Cognitive development in the preprimary years is fostered by hands-on manipulation of materials (Piaget, 1970; Shore, 1997). Children in settings with an inadequate number or variety of materials do not have as many opportunities to experiment and solve problems at their own pace.

Why free choice activities are positively related to language performance is less obvious. Analyses of what children actually did during free choice time showed that most often they chose to participate in physical and expressive activities and rarely elected to participate in preacademic activities (Lockhart, Xiang, & Montie, 2003). Two things likely to characterize free choice activities from a child's perspective are that the chosen activities are interesting and engaging and the difficulty level is suitable for the child. Perhaps when teachers propose specific activities instead of letting children choose, the activities are too easy or too difficult or simply not interesting for some children. Free choice activities provide the opportunity and, often, the necessity for children to interact verbally with other children in one-on-one or small group play--assigning roles for dramatic play, establishing rules for games, making plans for block building, and so forth. Further, the informal nature of free play provides an opportunity for teachers to engage children in conversation specific to their play and to introduce new vocabulary relevant to the children's interests, thereby promoting language acquisition. Research from the Home-School Study of Language and Literacy Development has shown positive links between both child-child interaction and pretend play during free play and children's performance on literacy measures (Dickinson, 2001a). Dickinson asserts that "free play is the time when children flex their linguistic and conceptual muscles and contribute to each other's development" (p. 253).

The opposite of this phenomenon may explain the negative relationship between teacher-proposed whole group activities and children's cognitive performance at age 7. By definition, activities proposed to the whole group are not tailored to each particular child's interest or learning ability. If an activity does not challenge a child or if it is too difficult, little useful learning takes place. In order to build cognitive skills, children need to solve problems and explore materials on their own. Learning and creativity take place when children confront situations that pique their interest and stretch their imaginations. A long tradition in the field of child development supports this constructivist view of early learning (e.g., Bruner, 1960; Piaget, 1970; Rogoff, 1990).

Findings that Vary Across Countries

Although even four findings that are common across all participating countries are enough to cast doubt on the belief that there are no universal relationships between preprimary practices and later behavior, the small number of them found also testifies to the importance of measured and unmeasured background factors that vary from country to country. A second set of findings presents relationships between level-2 setting predictors and children's age-7 performance that depend on the status of level-3 country predictors. It is important to note that

the country-level predictors are interpreted as proxies for unmeasured characteristics of countries, and we can only speculate as to how these characteristics relate to a country's culture and beliefs about children. Additionally, there were strong intercorrelations among some of the level-3 predictors. Therefore, any significant country-level predictor may reflect the covarying effects of a group of intercorrelated predictors and/or country characteristics that were not measured. Despite these caveats, the relationships involving significant country-level predictors provide clues to understanding the varying setting-level effects and hypotheses for further research.

The findings that vary across countries are presented below.

1. Increased adult-child interaction is related to better age-7 language scores in countries that have less adult-centered teaching or activities that require group response, and poorer language scores in countries that have more adult-centered teaching or activities that require group response.
2. Increased child-child interaction is related to better age-7 language scores in countries that have fewer whole group activities or more teachers who rank language skills among the most important, and poorer language scores in countries that have more whole group activities or fewer teachers that rank language skills among the most important.
3. Increased adult-child interaction is related to better age-7 cognitive performance in countries where teachers propose a lot of free choice activities, and poorer cognitive performance in countries where teachers propose few free choice activities.

The differential effects of adult-child and child-child interaction are intriguing; clearly, the nature of interaction differs in different situations. Reflection on the qualities of adult-child interaction in different classroom contexts provides clues to its meaning in the culture as a whole and its relationship with children's language performance. When adult-centered teaching is used, adults are primarily giving children information. Interaction is under the control of the adult and direction is from the adult to the child. When children respond as a group they generally repeat what is expected of them; responses require little thought and are apt to be automatic. The nature of child-child interaction changes according to the way children are grouped for activities. During large group activities there is little opportunity for children to have lengthy reciprocal conversations for the purpose of planning their play or solving problems. Social constructivist theory suggests that the most effective learning occurs when teaching is improvisational and children are allowed to interact and collaborate (Sawyer, 2004). In countries where adult-centered teaching, whole group activities, and group response are common, the prevailing cultural belief may be that it is of primary importance for children to listen, learn from, and obey those in authority. In contrast, in countries where child-centered teaching is typical or where children spend more time interacting with adults and other children individually and in small groups, children have a greater opportunity to express their own thoughts, opinions, and questions. In countries where these types of teaching practices are typical, adult-child interaction at home and at school is likely to encourage independent thought and freedom of expression, thus fostering language learning.

A similar phenomenon may explain the relationship among free choice activities, adult-child interaction, and children's cognitive performance. Although the study did not explore the nature of interaction during free choice activities, by definition, during free choice activities

adults do not direct children's play. In countries where free choice activities are more common, the prevailing cultural belief may be that children learn best from their own experiences and a major goal of adults' interaction with children is to enhance and expand those experiences.

It is surprising that group size was not found to be related to children's age-7 language or cognitive scores, because group size and adult-child ratio repeatedly emerge as important predictors of classroom quality and positive child outcomes in large-scale, multisite studies (e.g., Cost, Quality and Outcomes Study Team, 1995; NICHD Early Child Care Network, 1996; Ruopp et al., 1979); however, all the studies cited took place in the U.S. Examination of all the correlations of group size and adult-child ratio with observed process characteristics in each of the 15 countries that participated in Phase 2 of the study separately, showed that the number of positive associations was limited and inconsistent (Xiang, 2003), leading to the conclusion that relationships between group size and adult-child ratio and process characteristics are country-specific rather than universal. There is research evidence to support this view; for example, in Japan, having a large number of children in a classroom is considered desirable because it gives children more opportunity to learn from one another and reduces the demand for one-on-one interaction with the teachers (Tobin, Wu, & Davidson, 1991), an attitude that is in contrast to that in the U.S., where small group size is encouraged.

Limitations

Although the Preprimary Project is unprecedented in size and scope, in fact, a limited number of countries participated. Many cultures and regions were not represented (e.g., Africa and South America). Although the inclusion of country-level variables permitted exploration of reasons for varying effects across countries, these variables covary among themselves and with country characteristics that were not measured directly. We can only speculate about what cultural characteristics the country-level variables represent. As the unexplained variance at the setting-level indicates, many important setting-level influences were not measured. The observation instruments used, while relatively value-free and appropriate for use in many different countries, do not permit exact comparisons with other large scale studies of early childhood learning and quality of settings. The fact that age-4 developmental status measures were administered at different times during the school year means that any setting effects that took place prior to age-4 testing were adjusted for. In addition, the functional equivalence of the outcome measures limited the cross-country research questions that could be explored in the analysis.

Conclusion

The effect sizes found in this study are modest, but meaningful. The relationships between setting-level predictors and children's performance make intuitive sense and, in some cases, replicate findings from other studies. The findings from this study confirm that, despite the diversity of children's experiences in early childhood settings in different countries, there are at least some teaching practices and setting characteristics that lead to desirable outcomes across diverse countries. The implications that flow from the findings emphasize child-initiated activities and de-emphasize whole group instruction and are consistent with developmentally appropriate practices promoting active learning that have long been advocated by the National

Association for the Education of Young Children and others (European Commission, 1995; Head Start Bureau, 1984; Bredekamp, 1987).

Table 1
Achieved Sample for Phase 2 and Phase 3 by Type of Phase 2 Setting

Country/Region	Phase 2 *N* of settings	Phase 2 *N* of children	Phase 3 *N* of children	Retention rate (%)
Finland				
Child care centers	66	113	108	96
Greece (urban)				
Public kindergartens	49	91	63	69
Child care centers	22	66	31	47
Hong Kong (SAR)				
Kindergartens	26	104	88	85
Child care centers	25	100	86	86
Indonesia (Java)				
Urban kindergartens	39	152	123	81
Rural kindergartens	28	98	91	93
Ireland				
Nondisadvantaged preschools	25	86	81	94
Disadvantaged preschools	29	100	90	90
Disadvantaged national schools	27	100	94	94
Nondisadvantaged national schools	28	100	94	94
Italy				
Preprimary schools	143	501	246	49
Poland				
Urban preschools	75	292	252	86
Rural preschools	49	165	152	92
Spain (Catalonia)				
Public kindergartens	29	115	110	96
Private kindergartens	25	100	99	99
Thailand				
Educational programs	31	124	110	89
Child care centers	22	88	82	93
United States (6 sites)				
Head Start centers	31	109	59	54
Public school preschools	24	107	66	62
Other organized group programs	45	193	122	63
Total	838	2904	2247	

Table 2
Data Collection Instruments

Instrument	Description
Observation Systems	
Child Activities	Activities are coded into 1 of 12 categories: physical, expressive, storytelling/language, preacademic, religious, media-related, personal/social, expression of emotion, domestic, transitional, accidents, no active engagement. Interactions are noted.
Adult Behavior	Activities are coded into 1 of 8 categories: teaching, participation, nurturance, child-management, supervision, transitional, routine, and personal.
Management of Time	Observer keeps a running record of each activity proposed by the adult, the duration of the activity, and the proposed group structure (e.g., whole group, part group).
Questionnaires	
Family Background	Household composition, parents' educational attainment and employment, household income, and other characteristics.
Provider Survey	Physical characteristics of setting, management policies, teacher background, and materials available for children's use.
Teacher Expectations	Teachers rank 8 areas of development (language, motor, preacademic, self-assessment, self-expression, self-sufficiency, social skills with adults, social skills with children) in order of their importance for preprimary children to learn.

Table 3
Level 2 and Level 3 Predictors in the Hierarchical Linear Model

Variable name	Definition
Level-2 setting	
Predominant activity	MOT[a] activity most frequently proposed (free choice, physical/expressive, preacademic, or personal/social)[b]
Whole group	MOT percentage of proposed whole group activities
Adult-child interaction	CA[c] percentage of child interaction with adults
Child-child interaction	CA percentage of child interaction with peers
Group response	CA percentage of children's group response
No active engagement	CA percentage of children's no active engagement
Adult behavior	AB[d] typical behavior (teaching, participation, or no direct involvement with children)[b]
Adult listening	AB percentage of adult listening to children
Adult centered	AB percentage of adult-centered teaching
Mean parent education	Mean years of parents' education
Group size	# of children per class
Teachers' education	Number of full-time years of education
Materials	Variety of materials in setting
Level-3 country[e]	
Free choice_3	Mean % of free choice activity
Physical/expressive_3	Mean % of physical/expressive activity
Preacademic_3	Mean % of preacademic activity
Personal/social_3	Mean % of personal/social activity
Whole group_3	Mean % of whole group activities
Adult-child interaction_3	Mean % of child interaction with adults
Child-child interaction_3	Mean % of child interaction with peers
Group response_3	Mean % of children's group response
No active engagement_3	Mean % of children's no active engagement
Adult teaching_3	Mean % of adult teaching
Adult participation_3	Mean % of adult participation
Adult non-involvement_3	Mean % of adult noninvolvement
Adult listening_3	Mean % of adult listening to children
Adult-centered_3	Mean % of adult-centered teaching
Mean parent education_3	Country mean years of parents' education
Group size_3	Mean # of children per class
Teachers' education_3	Mean # of full time years of education completed
Materials_3	Mean # of varieties of materials in setting
Academic expectation_3	Mean beliefs about importance of academic dev
Language expectation_3	Mean beliefs about importance of language dev
Social expectation_3	Mean beliefs about importance of social dev

[a]MOT = Management of Time. [b]Dummy variable for category membership generated from a cluster analysis. [c]CA = Child Activities. [d]AB = Adult Behavior. [e]Based on data averaged across settings in a country.

Table 4
Number of Children and Settings Included in HLM Analysis by Country

	10-Country analysis		7-Country analysis	
Country	Settings	Children	Settings	Children
Finland	62	97	57	92
Greece	47	86	—	—
Hong Kong	50	167	45	150
Indonesia	67	211	—	—
Ireland	97	316	91	296
Italy	68	215	64	201
Poland	59	200	59	200
Spain	54	204	—	—
Thailand	50	181	50	181
U.S.	78	220	60	180
Total	632	1897	426	1300

Note. ' —' Not included in the sub-sample analysis due to unavailability of data on structural characteristics.

Table 5
Descriptive Statistics for Outcome and Predictor Variables

Variable	%	Mean	*SD*	Range
Outcome (*N* = 1897 children)				
Age-7 language scores (standardized)		0	1.00	-4.19 – 2.81
Age-7 cognitive scores (standardized)		0	1.00	-3.75 – 3.71
Level 1 (*N* = 1897 children)				
Age-4 language score (standardized)		0	1.00	-4.68 - 2.75
Age-4 cognitive score (standardized)		0	1.00	-2.24 – 3.20
Age at test (in months)		92.46	3.59	77.00 – 107.00
Age spread (in months)		37.32	3.13	26.00 – 53.00
Gender (male)	49%			
Number of siblings		1.22	1.12	0 – 14
Parents' education (in years)		11.38	3.59	0 – 24
Level 2 (*N* = 632 settings)				
MOT free choice[a]	25%			
MOT physical/expressive[a]	27%			
MOT preacademic[a]	25%			
MOT personal/social[a]	23%			
MOT whole group activity[b]		75.28%	23.96%	0 – 100%
CA adult-child interaction[b]		8.96%	11.68%	0 – 74.06%
CA child-child interaction[b]		27.15%	22.18%	0 – 95.42%
CA group response[b]		7.90%	10.98%	0 – 78.75%
CA no active engagement[b]		12.10%	8.93%	0 – 57.50%
AB teaching[c]	40%			
AB participation[c]	18%			
AB non-involvement[c]	43%			
AB listening[b]		2.70%	4.40%	0 – 40.00%
AB adult-centered teaching[b]		59.69%	27.76%	0 – 100%
Classroom average parent ed (in years)		11.55	2.93	3.88 – 22
Group size (*N* = 426)		20	8	4 - 49
Teachers' education (in years) (*N* = 426)		13.9	2.79	4 - 21
Materials[d] (*N* = 426)		49.27	15.27	1 - 80
Level 3[e] (*N* = 10 countries)				
MOT free choice_3		17.59%	7.50%	4.13 – 27.92%
MOT physical/expressive_3		24.66%	4.92%	17.61 – 32.43%
MOT preacademic_3		18.44%	6.09%	8.26 – 27.02%
MOT personal/social_3		17.17%	8.02%	10.46 – 36.65%
MOT whole group_3		76.05%	17.65%	41.05 – 94.33%
CA adult-child interaction_3		8.57%	7.24%	1.72 – 26.43%
CA child-child interaction_3		27.37%	17.35%	10.52 – 65.82%
CA group response_3		7.92%	5.21%	1.52 – 20.25%
CA no active engagement_3		12.19%	4.58%	6.18 – 19.13%
AB adult teaching_3		32.00%	7.58%	21.21 – 45.27%
AB adult participation_3		11.30%	6.69%	3.69 – 24.89%
AB adult non-involvement_3		20.60%	7.06%	11.38 – 32.34%

Variable	%	Mean	*SD*	Range
AB adult listening_3		2.80%	1.38%	0.89 – 4.87%
AB adult-centered teaching_3		60.51%	12.75%	43.60 – 87.30%
Academic expectation_3[f]		0.64	0.21	0.28 - 0.90
Language expectation_3[f]		1.01	0.25	0.68 – 1.39
Social expectation_3[f]		1.22	0.24	0.89 – 1.66

[a]Percentage of all settings for which 1 of 4 activities (free choice, physical/expressive, preacademic, or personal/social) was predominant in the setting. [b]Percentage of time spent in activity in setting. [c]Percentage of all settings for which 1 of 3 adult behaviors (teaching, participation, or non-involvement) was predominant. [d]Number of types of materials in a setting, derived from a list of 112 different types. [e]Computed across all settings in a country. Means are average percentages across 10 countries. [f]Mean level of importance based on teacher rankings from 0 to 2.

Table 6
Three-level HLM Estimates for Age-7 Language Outcomes

Fixed Effects	Estimates[a]	*SE*	Notes
Effects for Adjustment			
Age-4 language score (γ_{100})	0.30**	.02	Effects varied across countries
Adult participation_3(γ_{101})	0.07**	.02	
Age (γ_{200})	0.04	.03	
Age spread (γ_{300})	0.07*	.03	
Gender (γ_{400})	0.04	.03	
Siblings (γ_{500})	-0.08**	.02	
Parents' education (γ_{600})	0.19**	.03	
Mean parent education (γ_{010})	0.05	.03	
Effects of Interest			
Predominant activity			Free choice as reference
Physical/expressive (γ_{020})	-0.09	.07	
Preacademic (γ_{030})	-0.13^{+}	.07	
Personal/social (γ_{040})	-0.18**	.07	
Adult-child interaction (γ_{050})	-0.04	.04	Effects varied across countries
Group response_3 (γ_{051})	-0.11*	.04	
Adult-centered_3 (γ_{052})	-0.11*	.04	
Child-child interaction (γ_{060})	-0.05	.04	Effects varied across countries
Whole group_3 (γ_{061})	-0.17**	.05	
Language expectation_3 (γ_{062})	0.14**	.04	
Teachers' education (γ_{070})	0.07*	.07	Based on 7-country analysis

Random Effects	Unconditional	Unexplained	Explained
Child-level (e_{ijk})	.57092[b]	.46622	18.3%
Setting mean score (r_{0jk})--Overall	.24556**[b]	.14216**	42.1%
Country mean score (U_{00k})	.23139**[b]	NA	NA
Setting mean score (r_{0jk})--Portion available for setting-level predictors	.17148**[c]	.14216**	17.1%
Age-4 language score slope (U_{10k})	.00390**[d]	.00002	99.5%
Adult-child interaction slope (U_{05k})	.00709*[d]	.00002	99.7%
Child-child interaction slope (U_{06k})	.03076**[d]	.00006	99.8%

Table 7
Three-level HLM Estimates for Age-7 Cognitive Outcomes

Fixed Effects	Estimates[a]	*SE*	Notes
Effects for Adjustment			
Age-4 cognitive score (γ_{100})	0.32**	.03	Effects varied across countries
Adult teaching_3 (γ_{101})	-0.16**	.02	
Preacademic_3 (γ_{102})	0.13**	.03	
Age (γ_{200})	0.06*	.03	
Age spread (γ_{300})	0.11**	.03	
Gender (γ_{400})	0.02	.04	
Siblings (γ_{500})	-0.07**	.02	
Parents' education (γ_{600})	0.20**	.03	
Mean parent education (γ_{010})	0.03	.04	
Effects of Interest			
Whole group (γ_{020})	-0.10**	.04	
Adult-child interaction (γ_{030})	0.00	.03	Effects varied across countries
Free choice_3 (γ_{031})	0.18**	.05	
Materials (γ_{030})	0.09*	.04	Based on 7-country analysis

Random Effects	Unconditional	Unexplained	Explained
Child-level (e_{ijk})	.67284[b]	.57352	14.8%
Setting mean score (r_{0jk})--Overall	.26555**[b]	.17290**	34.9%
Country mean score (U_{00k})	.08387** [b]	NA	NA
Setting mean score (r_{0jk})--Portion available for setting-level predictors	.18963**[c]	.17290**	8.8%
Age-4 cognitive score slope (U_{10k})	.03537**[d]	.00002	99.5%
Adult-child interaction slope (U_{03k})	.00004*[d]	.00002	50.0%

[a]All numeric variables were standardized with a mean of 0 and standard deviation of 1. [b]Based on fully unconditional model. [c]After the level-1 predictors were included with grand mean centered, but before any level-2 predictors were included. [d]Before level-3 predictors were included in the corresponding model. NA: Not available because no random effect could be estimated due to inclusion of country dummy variables, which explained all the variance in countries' mean scores.
** $p < .01$; * $p < .05$.

References

Arnett, J. (1989). Caregivers in day-care centers: Does training matter? *Journal of Applied Developmental Psychology 10*, 541-552.

Bate, M., & Smith, M. (1978). *Manual of Assessment in Nursery Education*. Windsor, UK: NFER-Nelson Publishing Co.

Boehm, A. (1969). *Boehm Test of Basic Concepts*. New York: The Psychological Corporation.

Bracken, B. A. (1984). *Bracken Basic Concept Scale*. San Antonio, TX: The Psychological Corporation.

Bredekamp, S. (1987). Developmentally appropriate practice in early childhood programs serving children from birth through age 8. Washington, D.C.: National Association for the Education of Young Children.

Bruner, J. S. (1960). *The process of education*. Cambridge, MA: Harvard University Press.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park, CA: Sage Publications.

Campbell, F. A., Ramey C. T., Pungello, E. P., Sparling, J., & Miller-Johnson, S. (2002). Early childhood education: Young adult outcomes from the Abecedarian project. *Applied Developmental Science, 6*, 42-57.

Cost, Quality, and Outcomes Study Team. (1995). *Cost, quality, and child outcomes in child care centers, public report (2nd ed.)*. Denver: University of Colorado at Denver, Economics Department.

Cryer, D., Tietze, W., Burchinal, M., Leal, T. & Palacios, J. (1999). Predicting process quality from structural quality in preschool programs: A cross-country comparison. *Early Childhood Research Quarterly, 14*, 339-361.

Dickinson, D. K. (2001a). Large-group and free-play times. In D. K. Dickinson & P. O. Tabors (Eds.), *Beginning literacy with language* (pp. 223-255). Baltimore: Paul H. Brooks Publishing Co.

Dickinson, D. K. (2001b). Putting the pieces together. In D. K. Dickinson & P. O. Tabors (Eds.), *Beginning literacy with language* (pp. 223-255). Baltimore: Paul H. Brooks Publishing Co.

European Commission. (1995). *Preschool education in the European Union—Current thinking and provision.* Luxembourg: Office for Official Publication of the European Communities.

Feldhusen, J. F.; Houtz, J. C.; Ringenbach, S. (1972). Purdue Elementary Problem Solving Inventory. West Lafayett, IN.: Purdue University.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston, MA: Kluwer-Nijhoff.

Hammill, D. D., & Bryant, B. R. (1985). *Detroit Tests of Learning Aptitude—Primary*. Austin, TX.: PRO-ED..

Hammill, D. D., & Newcomer, P. L. (1982). Test of Language Development Intermediate. Austin, TX.: PRO-ED.

Harms, T., Clifford, R. M., & Cryer, D. (1998). *Early Childhood Environment Rating Scale* [Rev. ed.]. New York: Teachers College Press.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore: Paul K. Brooks Publishing Company.

Head Start Bureau. (1984). *Head Start Performance Standards* (DHHS Publication No. (ACF) 92-31131). Washington, D.C.: U.S. Department of Health and Human Services.

Hresko WP, Reid DK, Hammill DD. *The Test of Early Language Development*. Austin, Tex: Pro-Ed; 1981.

Iowa Tests of Basic Skills. (1972). Chicago: The Riverside Publishing Co.

Katz, L. G. (1968). *A study of changes in behavior of children enrolled in two types of Head Start classes.* Unpublished doctoral dissertation, Stanford University, Palo Alto, California.

Kaufman, A. S., & Kaufman, N. L. (1983). Kaufman Assesment Battery for Children. Circle Pines, MN.: American Guidance Service.

Kish, L. (1965. *Survey sampling*. New York: Wiley & Sons.

Kish, L. (1987). *Statistical design for research*. New York: Wiley & Sons.

Layzer, J., Goodson, B. D., & Moss, M. (1993). *Life in preschool: Volume one of an observational study of early childhood programs for disadvantaged four-year-olds*. Cambridge, MA: Abt Associates Inc.

Lockhart, S. D., Xiang, Z., Montie, J. (2003). Findings from concurrent observation systems. In D. P. Weikart, P. P. Olmsted, & J. Montie (Eds.). *A world of preschool experience: Observations in 15 countries* (pp. 162-196). Ypsilanti, MI: High/Scope Press.

Newborg, J., Stock, K. J., Wneck, L., Guidibaldi, J. & Svinicki, J. (1984). *Battelle Developmental Inventory*. Allen, Tx.: DLM Teaching Resources.

NICHD Early Child Care Research Network. (1996). Characteristics of infant child care: Factors contributing to positive caregiving. *Early Childhood Research Quarterly, 11,* 269-306.

Olmsted, P. P., & Weikart, D. P. (Eds.). (1994). *Families speak: Early childhood care and education in 11 countries.* Ypsilanti, MI: High/Scope Press.

Olmsted, P. P., & Montie, J. (Eds.). (2001). *Early childhood settings in 15 countries: What are their structural characteristics?* Ypsilanti, MI: High/Scope Press.

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

Piaget, J. (1970). Piaget's theory. In P. H. Mussen, (Ed.), *Carmichael's Manual of Child Psychology* (3rd ed., pp. 703-732). New York: John Wiley & Sons, Inc.

Reynolds, A. J., Temple, J. A., Robertson, D. L., & Mann, E. A. (2001). Long-term effects of an early childhood intervention on educational achievement and juvenile arrest: A 15-year follow-up of low-income children in public schools. *Journal of the American Medical Association, 285*, pp 2339-2346.

Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context*. New York: Oxford University Press.

Ruopp, R., Travers, J., Glantz, F., & Coelen, C. (1979). *Children at the center: Final report of the National Day Care Study, Vol 1.* Cambridge, MA: Abt Associates.

Sawyer, R. K. (2004). Creative teaching: Collaborative discussion as disciplined improvisation. *Educational Researcher, 33*, pp. 12-20.

Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (2005). Lifetime effects: The High/Scope Perry Preschool study through age 40. *Monographs of the High/Scope Educational Research Foundation, 14.* Ypsilanti, MI: High/Scope Press.

Shore, R. (1997). *Rethinking the brain: New insights into early development*. New York: Families and Work Institute.

Siraj-Blatchford, I., Sylva, K., Muttock, S., Gilden, R., & Bell, D. (2002). *Researching effective pedagogy in the early years* (Research Brief No. 356). Retrieved February 14, 2005 from http://www.dfes.gov.uk/research/data/uploadfiles/RB356.pdf

Sylva, K., Roy, C., & Painter, M. (1980). Oxford Preschool Research Project: Vol. 2 Childwatching at playgroup & nursery school. Ypsilanti, MI: High/Scope Press.

Tobin, J. J., Wu, D. Y., & Davidson, D. H. (1991). Forming groups. In B. Finkelstein, A. E. Imamura, & J. J. Tobin (Eds.). *Transcending stereotypes: Discovering Japanese culture and education* (pp. 109-117). Yarmouth, ME: Intercultural Press.

Vandell, D., & Wolfe, B. (2000, May). *Child care quality: Does it matter and does it need to be improved?* Retrieved June 14, 2000 from http://aspe.hhs.gov/hsp/ccquality00/ccqual.htm

Weikart, D. P. (Ed.). (1999). *What should young children learn? Teacher and parent views in 15 countries.* Ypsilanti, MI: High/Scope Press.

Weikart, D. P., Olmsted, P. P., & Montie, J. (Eds.). (2003). *A world of preschool experience: Observations in 15 countries.* Ypsilanti, Michigan: High/Scope Press.

Wolfe, E. W., & Manalo, J. R. (2003). *Scaling of the Preprimary Project Phase II and III Cognitive and Language Developmental Status Measures.* Unpublished manuscript available from the High/Scope Educational Research Foundation.

Woodcock, R. W. (1984). Woodcock Reading Mastery Test. Circle Pines, MN.: American Guidance Service.

Xiang, Z. (2003). Relationships between setting process and context. In D. P. Weikart, P. P. Olmsted, & J. Montie (Eds.). *A world of preschool experience: Observations in 15 countries* (pp. 197-226). Ypsilanti, MI: High/Scope Press.

Acknowledgements

The preparation of this manuscript was supported by the High/Scope Educational Research Foundation. The authors acknowledge the late Dr. David Weikart, who served as International Coordinator of the study, and the late Dr. Leslie Kish, who served as the international sampling referee. We are grateful to Dr. Patricia Olmsted, Jill Claxton, and Shannon Lockhart for their contributions to the project, to Dr. Edward Wolfe and Jonathon Manolo who evaluated the outcome measures, and to Dr. Yasuo Miyazaki, our consultant on hierarchical linear modeling. We would also like to acknowledge the country research coordinators: Mikko Ojala, Finland; Christos Frangos and Meni Tsitouridou, Greece; Sylvia Opper, Hong Kong; Diah Harianti, Indonesia; Nóirín Hayes, Ireland; Lucio Pusci and Egle Becchi, Italy; Malgorzata Karwowska-Struczyk, Poland; Mirea Montané and Dolors Iduarte, Spain; Nittaya Passornsiri and Pusadee Kutintara, Thailand.

Footnotes

[1]Leslie Kish, Ph.D., author of *Survey Sampling* (1965) and *Statisical Design for Research* (1987), served as international sampling referee supervising the developmental of each country's sampling plan.

[2]The items in the language developmental status measure were adapted from: the *Test of Early Language Development*, (Hresko, Reid, & Hammill, 1981).

[3]The items in the cognitive developmental status measure were adapted from: the *Manual for Assessment in Nursery Education.* (Bate & Smith, 1978).; the *Boehm Test of Basic Concept,.* (Boehm, 1969); and the *Bracken Basic Concept Scale,* (Bracken, 1984).

[4]Items for the cognitive developmental status measure were adapted from: the *Battelle Developmental Inventory,* (Newborg, Stock, Wnek, Guidibaldi, & Svinicki, 1984); the *BoehmTest of Basic Concepts,* (Boehm, 1969); the *Bracken Basic Concept Scale*, (Bracken, 1984); the *Detroit Tests of Learning Aptitude – Primary,* (Hammill & Bryant, 1985); the *Iowa Test of Basic Skills*, (1972); *the Kaufman Assessment Battery for Children,* (Kaufman & Kaufman, 1983); the *Purdue Elementary Problem Solving Inventory*, (Feldhausen, Houtz, & Ringenbach, 1972); and the *Woodcock Reading Mastery Test, (Woodcock, 1984).*

[5]Items for the language developmental status measure were adapted from: the *Test of Early Language Development*, (Hresko, Reid, & Hammill, 1981);the *Iowa Tests of Basic Skills* Level 7 Form 5 (1972);the *Battelle Developmental Inventory*, (Newborg, Stock, Wnek, Guidibaldi, & Svinicki, 1984); and the *Test of Language Development Intermediate*, (Hammill & Newcomer, 1982).