Early childhood education quality and child outcomes in China: Evidence from Zhejiang Province

Kejian Li[a], Yi Pan[b,†], Biying Hu[c], Margaret Burchinal[b], Allison De Marco[b], Xitao Fan[c], Jinliang Qin[a,*]

[a] *Hangzhou College of Preschool Teacher Education, Zhejiang Normal University, China*
[b] *Frank Porter Graham Child Development Institute, University of North Carolina at Chapel Hill, USA*
[c] *Faculty of Education, University of Macau, China*

ARTICLE INFO

Article history:
Received 2 September 2014
Received in revised form 11 January 2016
Accepted 16 January 2016
Available online 11 February 2016

Keywords:
Early childhood education
Quality
Child outcome
Association
China

ABSTRACT

Despite high rates of Chinese kindergarteners (3–6 years old) enrollment in early care and education (ECE), the quality of that care has not been widely examined. Following rapid economic growth in urban areas in the past three decades, there are growing concerns within China that families in urban and rural areas are experiencing an ECE opportunity gap. To address this concern, this study examined ECE quality and its association with child outcomes based on a relatively large sample of kindergartens in China. Using a stratified and random sampling method, the study recruited 1,012 children (age 3–6) from 178 classrooms in Zhejiang Province, a relatively developed region with a population of over 54 million people. We used the *Chinese Early Childhood Environment Rating Scale* to measure ECE quality and found moderately low quality for the study sample. Also, lower quality was observed in rural than urban areas, in private than public programs, and in programs with overall low parent education than those with high parent education. One dimension of quality, *teaching and interactions*, predicted child outcomes in language, early math, and social cognition as measured by the *Children's Developmental Scale of China (age 3–6)* in hierarchical linear models. The possible sociocultural and contextual reasons for these findings and implications for policymakers and practitioners are discussed in this paper.

1. Introduction

Nowadays, more than half of Chinese children aged 3–6 attend center-based childcare before primary school. As in many countries, kindergarten, called *'you'eryuan'* in Chinese, is the predominant form of center-based childcare for 3 to 6 year olds in China. In 2010, The Central Government of the People's Republic of China (2010) promulgated the *Compendium for China's Mid- and Long-Term Education Reform and Development*, which declared that 95% of Chinese children should receive at least one year of preschool education, while 75% of children should receive a three-year preschool education by 2020. According to National Bureau of Statistics of the People's Republic of China (2013), the proportion of 3–6-year-old children enrolled in childcare centers rose from 28% in 1991 to 62% in 2012. Thus, the focus on increasing access has raised questions about the quality of early care and education (ECE) in China. However, the quality of that care has not been widely examined. Furthermore, there are also concerns that different levels of ECE access and quality may exacerbate the growing opportunity gaps between urban and rural children (Hu, Zhou, Li, & Roberts, 2014), children attending public and private programs (Bu, 2008), and children of higher and lower socioeconomic statuses (SESs) (Zhang, Luo, Tao, Luo, & Dong, 2015). The purpose of this study was to describe the quality of ECE in a province of China and relate it to the early developmental outcomes of children.

1.1. Defining ECE quality in China

Under the influence of the deep-rooted cultural traditions of collectivism and respect for knowledge, Chinese ECE tended to focus on maintaining discipline, teaching children rules, basic art and academic skills (Li & Rao, 2005; Ng & Rao, 2005; Tobin, Hsueh, & Karasawa, 2009). Consistent with these traditional beliefs, most ECE involves large-group instructional and art/music activities and

* Corresponding author at: Hangzhou College of Preschool Teacher Education, Zhejiang Normal University, 310012, China Fax: +86 57181952029.
E-mail address: qjl@zjnu.edu.cn (J. Qin).

† Deceased (2016). In memory of Dr. Yi Pan who cared deeply about the provision of quality early childhood education for disadvantaged children in China. The research team, greatly saddened by the loss of a friend and colleague, will carry on his dream of improving the lives of children in poverty in China.

http://dx.doi.org/10.1016/j.ecresq.2016.01.009

group responses to teachers, although direct instruction in literacy and calculating skills is forbidden during the kindergarten years (Zhu & Wang, 2005). Over the past three decades, Chinese ECE professionals have introduced some Western ECE quality concepts into Chinese ECE regulations and practices (Hu, 2014; Zhu & Zhang, 2008), such as play-based curriculum and child-centered teaching approaches (Ministry of Education of the People's Republic of China, 2001). However, a recent national survey on children's daily activities in 440 classrooms from 11 provinces indicated that kindergarteners (aged 3–6) spent 46% of their time on whole-group activities during the observation day and only 19% on free play in interest-center activities (Liu, 2011).

A few studies have explored the quality of Chinese kindergarten education (Liu, 2011; Pan, Liu, & Lau, 2010; Xiang & Liao, 1995). In the early 1990s, Xiang and Liao (1995) reported relatively poor ECE quality in a nationwide study in China that included over 400 ECE programs from six provinces. Their observations indicated that Chinese ECE programs lacked age-appropriate materials for children's learning and play and provided few opportunities for free play, individualized interactions with teacher(s), and social interactions with peers (Xiang & Liao, 1995). Chinese kindergarten teachers tended to focus on academic skills, discipline and rule enforcement rather than socio-emotional development (Xiang, 1995). A recent study of 28 classrooms from 14 kindergartens in Beijing obtained largely consistent findings using an unpublished measure similar to *Early Childhood Environment Rating Scale (ECERS)* and reported that these programs were of low to moderate quality (Pan et al., 2010). The study indicated that in most of the observed classrooms, teacher-directed group activities occupied most of a school day and teachers provided little time and few opportunities for free play. Most teacher–child interactions were one-way in a large-group manner and teachers paid little attention to children's individual learning needs and interests (Pan et al., 2010).

ECE quality in Chinese kindergartens may vary along several dimensions. Due to the long-lasting urban/rural divide in China, the rapid economic growth in urban areas in the past three decades has led to an upward social mobility for many urban families, but relatively stagnant social mobility for many families in rural areas that have not seen such growth (Guo, 2008; Pang, 2006; Tang & Zhao, 2008). Then, there are concerns that children in urban areas enjoy more opportunities than children in rural areas, including access to higher quality ECE (Wu, Yong, & Cai, 2012). Lacking of sufficient funding and qualified teachers, ECE programs in rural areas are found to have much lower quality on average than those in urban areas (Hu, Zhou et al., 2014). By auspice, Chinese kindergartens can be largely divided into public and private kindergartens. According to statistics by Ministry of Education of the People's Republic of China (2013), private kindergartens take up 69% of all ECE programs in China. Like the U.S., many Chinese private kindergartens are for-profit. However, in contrast to the U.S., the majority of Chinese private programs largely serve children from mid- to low-income families (Bu, 2008). Although both public and private kindergartens charge tuition, public kindergartens receive significantly more public funding and private kindergartens operate mostly based on private funding sources (e.g., student tuitions). Thus, public programs tend to be better funded and better equipped and have more qualified and stable teacher forces than private ones (Hu & Li, 2012). Meanwhile, the quality of kindergartens might differ dramatically based on the SESs of the families they serve. In China, like in many other countries, more-advantaged children might have better chances to attend higher quality kindergartens than less-advantaged children (Zhang et al., 2015). Furthermore, one can assume that these dimensions may not be independent and these divides may have impacts, possibly in an interactive manner, on child developmental outcomes, as more-advantaged children are more likely to live in urban areas and attend high quality public programs.

1.2. Current research relating ECE quality to child outcomes

1.2.1. ECE quality and child outcomes

In many U.S. studies, ECE quality has been related to modest but statistically significant gains in language, cognitive, and social skills (Burchinal, Magnuson, Powell, & Hong, 2015). Evidence from experimental and rigorous quasi-experimental studies has demonstrated that quality ECE has modest to moderate causal effects on child outcomes (Yoshikawa et al., 2013), and evidence from large-scale observational studies has shown positive yet modest associations between child care quality and child outcomes (Burchinal, Kainz, & Cai, 2011; Keys et al., 2013). Very modest causal associations were detected in a recent study using instrumental variables (Auger, Farkas, Burchinal, Duncan, & Vandell, 2014), suggesting gains in language and math related to global process quality measures. Similar associations between global quality and child outcomes have also been reported in developing countries such as India (Rao, 2010), Bangladesh (Aboud, 2006; Aboud & Hossain, 2011; Moore, Akhter, & Aboud, 2008), Kenya, Uganda, Tanzania (Malmberg, Mwaura, & Sylva, 2011), Jamaica (Baker-Henningham, Walker, Powell, & Gardner, 2009), Costa Rica (San Francisco, Arias, Villers, & Snow, 2006), and Chile (Leyva et al., 2015). There is some evidence that children from less-advantaged backgrounds may benefit more when they experience higher quality childcare than children from more-advantaged backgrounds in both the U.S. (Burchinal et al., 2015) and around the world (Engle et al., 2011; Heckman, 2010).

One potential reason for the relatively modest associations between ECE quality and child outcomes is that ECE quality is often defined and measured globally and measures of more specific aspects of quality may provide better predictions (Bryant, Zaslow, & Burchinal, 2010; Peisner-Feinberg, & Yazejian, 2010; Zaslow, Martinez-Beck, Tout, & Halle, 2011). Recent meta-analysis of data from seven large-scale child care studies indicated stronger associations between child outcomes and ECE quality when quality was measured more specifically than globally, as measured by the *Early Childhood Environmental Rating Scale-Revised* (ECERS-R) total score (Burchinal et al., 2011). The studies found stronger associations between specific dimensions of ECE quality and aligned domains of child outcomes (Burchinal et al., 2015). For example, it was found that the "*interaction*" factor encompassing items related to teacher/peer interactions and teaching children in a responsive manner modestly predicted children's language and social skills, but ECERS-R total score did not (Burchinal et al., 2015). Other findings looked at specific quality measures from other ECE quality instruments (e.g., CLASS) and reported stronger associations with child outcomes for these specific quality measures than for the global ECERS-R total scores (Burchinal et al., 2015).

However, in a large-sample study, a research team reported mixed results for the associations between several ECERS-R quality dimensions (total score and three factor scores) and child outcomes (cognitive, social-emotional, and health) (Gordon, Fujimoto, Kaestner, Korenman, & Abner, 2013). They found that there were no significant associations between the ECERS-R total and factor scores and child outcomes in reading and math. A few significant associations were found between children's socio-emotional and health outcomes, but they were small and not in a pattern consistent with domain specificity (Gordon et al., 2013). Although results were not consistent across studies, it is reasonable to conclude that the strength of associations between ECE quality and child outcomes are stronger when specific dimensions of quality are more closely aligned with certain domains of child outcomes.

1.2.2. Relating ECE quality to child outcomes in China

To date, very few large scale studies have closely examined the overall quality of ECE and its associations with child outcomes in China despite the government's goal of ensuring both ECE access and quality for Chinese children aged 3–6 (The State Council of the People's Republic of China, 2010). Based on a relatively very small sample (10 classrooms), Rao, Sun, Zhou, and Zhang (2012) examined the relationships between types of ECE experience in a rural county of Guizhou Province. Classroom quality was evaluated on several widely-used measures, including the ECERS-R (Harms, Clifford, & Cryer, 1998), the *Early Childhood Environment Rating Scale-Extension* (ECERS-E) (Sylva, Siraj-Blatchford, & Taggart, 2003), and the *Early Childhood Classroom Observation Measure* (ECCOM) (Stipek & Byler, 2004). It was found that the only one kindergarten classroom had higher quality than the other nine classrooms, and that children attending kindergarten had higher school readiness scores than those who attended pre-primary classes, Grade One classrooms of primary school, or had no ECE experience (Rao et al., 2012). This study provided some support that higher ECE quality was associated with better child outcomes. However, the researchers did not aim to systematically examine the association between ECE quality and child outcomes. Furthermore, the small and homogenous rural sample did not permit more in-depth examinations, such as examinations of the differential ECE quality across different auspices and geographic locations or differential effects of ECE quality on child outcomes in different domains.

Despite the large number of Chinese children in kindergarten settings and some evidence that kindergarten quality may be low in China, very little is known about kindergarten quality and its relationship with child outcomes in the Chinese context. Although previous studies have provided much-needed information, they have been relatively small and may not represent what is typical in China. For example, Rao et al. (2012) conducted their study in a single county of an inland underdeveloped province located far from the eastern seaboard, where a great deal of economic growth had occurred over the past three decades. Pan et al. (2010) examined only 14 kindergartens' quality in Beijing, the capital city and one of the most prosperous metropolitan areas in China, but not a typical province. And the tendency of these studies to use convenient and/or very small samples may prevent them from describing quality with much generalizability.

ECE is considered an important tool in addressing opportunity gaps around the world (Engle et al., 2011; Heckman, 2010), and China appears to view access to quality ECE as a means of addressing the growing gap between the economically vibrant urban areas and economically depressed rural areas (The State Council of the People's Republic of China, 2010). It is hoped that high quality ECE may serve as a protective factor for more-disadvantaged children by promoting their development and decreasing the achievement gap between them and their more-advantaged peers. To achieve this goal, it is important to understand more about the quality of ECE settings and its associations with child outcomes in a larger and more representative sample than previously examined. The current study attempted to fill this research gap.

1.3. The current study

Based on a large-scale kindergarten sample from an eastern province in China, the current study aimed to examine ECE quality and its associations with child developmental outcomes so as to inform stakeholders of how to better support and enhance children's early learning experiences. It attempted to answer the following research questions.

1. What is the ECE quality in this province, and does it vary by auspice (public/private), type of community/location (urban/rural), and center-level parental education (high/low) ?
2. To what extent does ECE quality predict child outcomes in language, early math, and social cognition?
3. Is ECE quality a stronger predictor of child outcomes for less-advantaged children (e.g., children in rural areas, children with less-educated parents)?

This study was based on the Chinese context and prior empirical ECE research. We expected that ECE quality would be higher in public rather than private programs, in urban rather than rural settings, and in kindergartens with overall high parent education rather than those with low parent education. Based on prior research findings, we anticipated that ECE quality would be related to attending children's language, early math, and social cognition skills and that those associations would be stronger for children in rural areas or with less-educated parents than their more advantaged peers. However, considering the complexity of child developmental process from the ecological perspective (Bronfenbrenner, 1979; Bronfenbrenner & Morris, 1998) and findings from other countries, we anticipated that these associations would be modest in magnitude.

2. Method

2.1. Participants

The participants included 1012 children randomly sampled from 178 classrooms in 91 Chinese kindergartens in Zhejiang Province, located on China's central eastern coastline with a population totaling over 54 million people. This province is one of the most developed areas in China and has both urban and rural areas. The 91 participating kindergartens were selected to represent three stratification factors: (1) economic development level of the area where the kindergarten was located (41, 21, and 29 kindergartens from areas of high, medium, and low level of economic development respectively); (2) ECE program quality classifications based on the current provincial government rating system (17 high, 32 medium, and 42 low rating kindergartens); (3) location (33 urban and 58 rural kindergartens). The selection criteria ensured that there were sufficient numbers of centers in each category according to the ECE contexts in the province. The 91 kindergartens ranged from 86 to 946 students in size. Their monthly tuitions varied from RMB100 to RMB2500 (roughly USD$16 to USD$400), varying as a function of program type, location, and governmental rating. All of these kindergartens were full-time programs that operated about 8 h daily.

Next, the researchers randomly selected two classrooms with different age group populations from each kindergarten on average. The 178 classrooms selected included 45 classrooms with 3-to-4-year-old students, 53 classrooms with 4-to-5-year-old students, 74 classrooms with 5-to-6-year-old students, and 6 mixed-age classrooms. The quality of these classrooms was observed and recorded. Six children (three boys and three girls) were randomly sampled from each classroom for developmental assessment in the areas of language, early math, social cognition, and motor skills. Complete child developmental outcome data were collected for 1012 children.

2.2. Measures

2.2.1. Classroom quality observational measure

The *Chinese Early Childhood Environment Rating Scale (trial version)* (CECERS; Li & Hu, 2012), a newly developed quality mea-

surement tool, was used to measure classroom quality. Modeled roughly after the ECERS-R (Harms et al., 1998), the CECERS uses a 7-point scoring system ranging from 1 (inadequate) to 3 (least acceptable), 5 (good), and 7 (excellent) to measure the quality of Chinese ECE programs for children aged 3–6. The CECERS comprises 51 items organized in eight subscales: (1) *Space and Furnishings* (nine items); (2) *Personal Care Routines* (six items); (3) *Curriculum Planning and Implementation* (five items); (4) *Whole-Group Instruction* (seven items); (5) *Activities* (nine items); (6) *Language-Reasoning* (four items); (7) *Guidance and Interaction* (five items); and (8) *Parents and Staff* (six items).

It should be noted that heavily substantive adaptations of the ECERS-R were made based on Chinese ECE contexts so as to improve the cultural adaptability of the measure. More importantly, the authors of CECERS created the whole new subscale of *Whole-Group Instruction* to reflect and evaluate the process quality of group teaching activities (Li, Hu, Pan, Qin, & Fan, 2014), which are prevalent in Chinese kindergartens and much valued by Chinese ECE practitioners (Li & Rao, 2005). An indicator-level content analysis showed that 53.4% of the total 685 indicators of CECERS were adapted, to different extent, from the ECERS-R while the other 46.6% were newly created (Li et al., 2014). Based on a national wide survey on 176 expert practitioners, the CECERS was considered as a well designed, much needed and culturally appropriate ECE quality measurement tool by the overwhelming majority of the reviewers (over 90%) (Hu, Vong, Chen, & Li, 2014). Therefore, the CECERS can be comparable to but not equal to the ECERS-R.

Results from a recent validation study indicated that the CECERS had good psychometric properties (Li et al., 2014). Cronbach's alphas at the subscale level ranged from 0.83 to 0.93 with the total scale at 0.96. And the item-level inter-rater reliability estimates ranged from 0.70 to 0.90 (overall mean of 0.77) with subscale-level inter-rater reliability estimates ranging from 0.85 to 0.94 (with the mean of 0.89), which indicated good inter-rater reliability. Accumulated evidence indicated the CECERS a reliable, valid and culturally appropriate instrument for measuring the environment and process quality of Chinese ECE programs (Li et al., 2014).

The structure of the CECERS was examined using exploratory factor analysis (EFA) with oblique rotation. Two factors were extracted based on multiple criteria, e.g., the parallel analysis result and scree plot (Li et al., 2014). The results indicated that two factors were highly correlated and accounted for 61% of the variance ($r = .76$). The first factor labeled *Provisions for Learning* included 21 items (most items come from subscale of *Space and Furnishings* and subscale of *Activities*) that measured the ECE program's provision of spaces, facilities, and age-appropriate materials and activities in support of children's learning. The second factor labeled *Teaching and Interactions* included 13 items (6 items come from subscale of *Whole-Group Instruction,* and 5 items come from subscale of *Guidance and Interaction* with the other 2 items from subscale of *Language-Reasoning* subscale) that measured the use of developmentally appropriate whole-group instruction, care giving, and interactions between children, teacher(s), and peers (see online Supplementary materials for details of the factor structure of the CECERS).

2.2.2. Child developmental measure

The *Child Developmental Scale of China (age 3–6)* (CDSC; Zhang, Zhou, Chen, Zhao, & Wang, 1992) was administered to collect the child outcome data. Designed to assess the overall development of children aged 3–6, the CDSC is a norm-referenced test that contains four subscales: (1) *Language* (25 items), evaluation of a child's vocabulary, understanding, and expression/use of language; (2) *Early Math* (56 items), evaluation of a child's classifying, shape cognition, reasoning, numbering, and calculating abilities; (3) *Social Cognition* (25 items), evaluation of a child's basic social knowledge, perception of social relationships, and early moral judgment; and (4) *Physical Movement* (five items), evaluation of a child's gross motor and fine motor skills. The authors reported that the internal consistency of the CDSC ranged from 0.71 to 0.95 with a test–retest reliability of 0.89 (Zhou & Zhang, 1994). The scale also exhibited good content, construct, concurrent validity (CDSC test scores significantly correlated with scores on the Stanford–Binet Intelligence Scale for children in China, $r = .603$), and criterion validity (children's CDSC test scores highly correlated with their school achievement test scores, $r = .740$). This study focused on the CDSC measures of *Language*, *Early Math*, and *Social Cognition* (see online Supplementary materials for detailed introduction of the three subscales of the CDSC), and we chose to use raw scores rather than normed scores due to the concern that norms were outdated.

2.2.3. Kindergarten director and teacher questionnaires

The CECERS evaluators asked the participating directors and the lead teacher of sample classrooms to complete a questionnaire to retrieve detailed information about the kindergartens, classrooms, and teachers. At the kindergarten level, information was collected on location (urban/rural), size, auspice (public/private), quality rating by the current government rating system, number of teachers, teacher incomes, and financial information (e.g., yearly public funds, tuitions, and expenditures). At the classroom/teacher level, information was collected on age group, class size, number of staff members in the classroom, teacher's age, years of teaching experience, teacher education level, teaching certificate, professional rank, and tenure status, etc.

2.2.4. Family demographic information

Parents of the recruited children were asked to provide basic family background information on parent age, occupation, education level, and yearly/monthly income. There were high levels of missing data on parent occupation and family income, possibly due to the parents' unwillingness to reveal sensitive information about their households. Therefore, we decided not to use parents' occupation and income as variables in our study.

2.3. Data collection

The researchers trained 28 research assistants (all ECE graduate students) as CECERS raters. The training consisted of four days of intensive study of the instrument and five days of field rating practice. Trainees were randomly assigned to trainers each day (two to three trainees for every trainer) for the field rating practice. Each person completed ratings on the CECERS and then examined the agreement for each item/sub-item with the trainer at the end of the field-training day. Quality rating techniques and the issues related to agreement for each item/sub-item were discussed. A certification test was then conducted. To pass the certification, each trainee was required to demonstrate a median level of agreement with the trainer's observation results as measured by a weighted kappa of 0.6 or higher. The exact agreement proportion was 0.63 on average, and the within-1-point agreement was 0.91. The weighted kappa inter-rater reliabilities were 0.6 or higher for almost all of the items, indicating substantial to outstanding inter-rater agreement. The researchers also trained 13 ECE graduate students to be CDSC raters. The training included 1 day of intensive study on the CDSC and 5 days of field-testing that included independent testing and a group discussion of the results. After 5 field-testing days, all 13 of the raters demonstrated their ability to administer the CDSC reliably (Cronbach's $\alpha = 0.9$ for the whole scale). The study followed the typical procedures for obtaining informed consent from the kindergartens and from the parents of the children involved in the assessment.

In spring 2012, two raters from the research team visited each of the 178 classrooms to conduct classroom observation. Each observation lasted about 6 h (typically 4 h in the morning and 2 h in the afternoon). The two raters observed and scored each classroom independently to account, in part, for rater variability (Chen, Hu, Fan, & Li, 2014). After completing their observations and scoring, they discussed their ratings and reached consensus scores on all of the items. This practice proved to be effective in significantly improving the reliability of the measurement (Chen et al., 2014). In addition, a half-hour interview was conducted with the classroom teacher during the children's naptime to gather information that was not directly observable. During the classroom observation, another trained rater administered CDSC on six children from the classroom. Each test took place in a separate and quiet room. Assessment time ranged from 20 to 30 min.

The evaluators gave the questionnaires to center directors and classroom teachers at the beginning of their observation days and retrieved the completed questionnaires when they left at the end of observation. Family information of the participating children was collected by reviewing the families' information registered at the time of their children's enrollment and/or teacher's telephone interviews with parent(s).

2.4. Analytic plan

Analysis proceeded in five steps. First, we excluded children who spent fewer than 6 months at their centers at the time of data collection with the rationale that the exposure might not be sufficient to have an observable effect on their development. As a result, 77 out of the 1012 observations were excluded, and the data for 935 children from 163 classrooms in 88 kindergartens were used in final analyses. Second, based on the recent EFA results for CECERS (Li et al., 2014), factor scores were calculated by averaging the scores of the items that loaded on each factor.

Third, missing data were addressed using multiple imputation (MI) techniques. Researchers typically use model-based MI under the assumption that data are missing at random (Kenward & Carpenter, 2008; Little & Rubin, 2002) to address missing data. In this study, an imputation model was constructed to account for data clustering and to incorporate basic family characteristics in addition to the variables used in final regression analysis. Imputed values of missing variables were obtained by drawing multiple independent samples from the modeled distributions of corresponding variables given all of the other auxiliary variables to replace the missing data and form multiple complete datasets. Ten imputation datasets were produced. All analyses were conducted on each dataset, and parameter estimates were combined across all of the datasets to produce final estimates (Rubin, 1987). Multi-level MI was conducted using REALCOM-IMPUTE (Goldstein, 2009) and Stata 12® (StataCorp, 2013).

Fourth, three-level hierarchical linear model (HLM) analysis was conducted with the observed classroom quality variables (i.e., the two factor scores of the CECERS) as the predictors of interest to investigate whether they were associated with child outcomes accounting for data clustering. The researchers ran a set of three-level HLMs for each child outcome. Model 1 consisted of the main effect of two CECERS factor scores, child age and gender, time exposed to ECE, mother's years of education, whether the kindergarten was public, and whether the kindergarten was located in an urban area (Eq. (1)). Model 2 added interaction terms between the CECERS factor scores and whether both of the child's parents had education levels lower than high school. Model 3 added interaction terms between the CECERS factor scores and whether the kindergarten was located in urban areas. Subgroup analysis using the Model 1 covariates was then conducted for the children whose parents had education levels lower than high school and otherwise and for the children from urban and rural areas. HLMs were conducted in SAS 9.3® (SAS Institute, 2011).

$$y_{ijk} = \beta_0 + \beta_1 \times \text{Provision}_{jk} + \beta_2 \times \text{Teaching}_{jk} + \beta_3 \times \text{Age}_{ijk} + \beta_4 \times \text{Male}_{ijk} + \beta_5 \times \text{ECE.Exp}_{ijk} + \beta_6 \times \text{Mom.Ed}_{ijk} + \beta_7 \times \text{Public}_k + \beta_8 \times \text{Urban}_k + \mu_{00k} + \nu_{0jk} + \varepsilon_{ijk} \quad (1)$$

Finally, a series of sensitivity analyses were conducted to confirm the robustness of the results. The same analytic models were fitted to the complete study sample data, including the 77 children who had less than six months of experience in their current childcare setting. The results were consistent with the results presented in this paper (see online Supplementary materials for results of sensitivity analyses with the full sample).

3. Results

3.1. ECE quality and variations

Descriptive analyses were conducted to describe ECE quality in Zhejiang Province, China. As shown in Table 1, the kindergarten sample had an average of nine classrooms (range of 3–25), 19 lead teachers (range of 3–51), five assistant teachers (range of 0–19), and 291 children (range of 91–946). The average class size was 35 (range of 11–63), with a mean child/adult ratio of 18.2 (range of 3–46). On average, the lead teachers had 9.3 years of teaching experience and 16.5 years of education. Of all the classroom lead teachers, 38% had a BA degree and 77% had a teaching certificate. As to the observed classroom quality, the CECERS scores suggested moderately low quality on average, with means of 3.3 overall (SD = 1.0), 3.0 (SD = 1.0) for the *Provisions for Learning* score and 3.5 (SD = 1.0) for the *Teaching and Interactions* score.

We then tested the extent to which ECE quality varied by auspice (pubic versus private), type of community/location (urban versus rural), and parent education level (high versus low). Results indicated that clear gaps existed between these comparisons on both structural quality and observed classroom quality (see online Supplementary materials for detailed analysis results).

3.1.1. ECE quality by auspice

Compared to private kindergartens, public kindergartens had more teachers, classrooms, children, expenditures, and classrooms were rated as being higher quality on the CECERS.

3.1.2. ECE quality by type of community/location

Compared to rural kindergartens, urban kindergartens had more teachers, aids, classrooms and expenditures, and classrooms were rated as being higher quality on the CECERS.

3.1.3. ECE quality by parent education level

Compared to kindergartens with overall low parent education, kindergartens with overall high parent education had more teachers, aids, classrooms, children, space and expenditures, and classrooms were rated as being higher quality on the CECERS.

A cross-tabulation of kindergartens by location, auspice, and overall parent education level showed the tendency of families with higher and lower parent education who used public or private childcare in urban and rural communities. As shown in Table 2, results indicated that the proportion of kindergartens with high parent education levels was significantly higher in urban areas than in rural areas (93% versus 28%, $\chi^2(1) = 34.45$, $p < .001$). Meanwhile, kindergartens with low parent education are more likely to be found in rural areas than in urban areas (72% versus 7%,

Table 1
Center and classroom characteristics: descriptive statistics.

	N	Mean	SD	Range
Center characteristics (*N* = 88)				
Number of teachers	73	16.85	10.43	3–51
Number of teacher aids	70	5.45	4.81	0–19
Number of classrooms	75	8.68	4.33	3–25
Number of children	72	290.95	158.52	91–946
Size of activity room area per child (m^2)	71	4.98	12.16	1–70
Size of nap area per child (m^2)	71	2.21	3.89	0–28.7
Size of outdoor activity area per child (m^2)	69	4.58	3.21	0.32–16.1
Center financing in U.S. Dollars (*N* = 88)				
Yearly expenditure on employees (in $1000s)	67	18.25	20.46	2–97
Yearly government funds received (in $1000s)	63	137.94	265.82	0–1158
Yearly expenditure/cost per child	76	793.23	622.71	88–2954
Yearly tuition per child	74	428.83	314	19–2049
Average teacher monthly salary	71	398.28	171.52	0–801
Average aid monthly salary	64	210.62	60.61	0–377
Structural classroom quality (*N* = 163)				
Enrolled class size	161	35.45	7.80	11–63
Observed class size	160	33.39	8.00	11–54
Child-to-teacher ratio	158	18.23	8.59	3–46
Teacher's age[a]	158	31.08	6.04	22–49
Teacher's experience (years)	157	9.29	6.37	0.5–30
Teacher's education (years)	160	16.50	2.42	9–19
Teacher with BA degree (%)	160	38%		
Teacher with teaching certification (%)	159	77%		
Teacher with professional rank (%)	163	56%		
Teacher under permanent contract (%)	152	34%		
Observed classroom quality (*N* = 163)				
CECERS–provisions for learning score	163	2.89	0.95	1.19–5.00
CECERS–teaching and interactions score	163	3.48	0.95	1.15–6.54
CECERS–total score	163	3.29	0.93	1.35–5.86

[a] At classroom level, 'teacher' refers to the primary teacher of the classroom.

Table 2
Cross tabulation of childcare centers by location, parent education level and auspice.

		High parent education (*N* = 45)			Low parent education (*N* = 43)		
	N (%)	Public (*N* = 24)	Private (*N* = 21)	Sum	Public (*N* = 18)	Private (*N* = 25)	Sum
Childcare center (*N* = 88)	Urban (*N* = 31)	14 (45%)	15 (48%)	29 (93%)	0 (0%)	2 (7%)	2 (7%)
	Rural (*N* = 57)	10 (18%)	6 (10%)	16 (28%)	18 (32%)	23 (40%)	41 (72%)

Note: At center level, high and low parent education levels are defined in terms of whether the overall percentage both parents with education levels lower than high school is below or above 30% (site sample median).

$\chi^2(1) = 35.37$, $p < .001$). Furthermore, rural private kindergartens are more likely to have low parent education than high parent education (40% versus 10%, $\chi^2(1) = 9.97$, $p < .01$). Therefore, the rural and private kindergartens, which had the higher proportions of low parent education, were the most disadvantaged according to many measures of structural quality and observed quality on the CECERS.

3.2. Relating ECE quality to child outcomes

The next set of analyses tested the extent to which child outcomes in language, early math, and social cognition were related to ECE quality. Table 3 shows the descriptive statistics for the measures of the children and their families, including children's raw scores on CDSC measures of language, early math, and social cognition (see online Supplementary materials for detailed information on child outcomes of different age groups). The descriptive statistics indicated that the children sample was 5.3 years old on average (range of 3–6.6), 50% of them were boys, 5% belonged to ethnic minority groups, and they averagely had 24.6 months of childcare experience in center (SD = 11.24). On average, their parents had 11.6–11.9 years education (SD = 3), and 33% of their both parents' education was below high school.

Table 4 shows the correlations between the CECERS scores and structural quality variables and child outcomes on CDSC including language, early math, and social cognition. As shown in Table 4, the three CECERS scores (two factor scores and the total score) are highly correlated, and they are consistently correlated with the three child outcome measures (i.e., CDSC language, early math, and social cognition). Among classroom-teacher variables, child-to-teacher ratio and teacher's education significantly and consistently correlated with the three CECERS scores, while teacher's experience shows relatively low correlations with CECERS *Teaching and Interactions* score and total score.

Table 3
Child and family characteristics and child outcomes: descriptive statistics.

	N	Mean	SD	Range
Child characteristics (*N* = 935)				
Age (years)	935	5.30	0.81	3–6.6
Male (%)	935	50%		
Racial/ethnic minority (%)	935	5%		
Time in center (months)	935	24.64	11.24	6–72
Family characteristics (*N* = 935)				
Mother's age (years)	846	33.52	4.27	23–50
Mother's education (years)	874	11.61	3.35	3–22
Father's age (years)	847	35.81	4.86	24–58
Father's education (years)	895	11.94	3.18	6–22
Both parents' Ed. below HS	890	33%		
Child outcomes on CDSC (*N* = 935)				
Language total score	935	20.97	5.04	6–32
Early math total score	935	33.14	8.80	5–57
Social cognition total score	935	16.94	3.91	3–29

Table 4
Correlations between CECERS quality measures and structural quality and child outcomes.

	CECERS provisions	CECERS interactions	CECERS total
Classroom characteristics ($N = 163$)			
Observed quality: CECERS			
Provisions for learning		0.77***	0.95***
Teaching and interactions			0.92***
Total score			
Structural classroom quality ($N = 163$)			
Enrolled class size	−0.04	−0.11	−0.06
Observed class size	−0.06	−0.14	−0.09
Child-to-teacher ratio	−0.45***	−0.45***	−0.49***
Teacher's experience	0.15	0.18*	0.17*
Teacher's education	0.45***	0.45***	0.48***
Child outcomes ($N = 935$)			
CDSC language	0.17***	0.25***	0.23***
CDSC early math	0.18***	0.26***	0.23***
CDSC social cognition	0.20***	0.25***	0.24***

* $p < .05$.
*** $p < .001$.

Table 5
HLM analyses: testing associations between childcare quality and child outcomes.

	CDSC language			CDSC early math			CDSC social cognition		
	B	(SE)	ES	B	(SE)	ES	B	(SE)	ES
Model 1									
CECERS Scores									
Provisions for learning	−0.30	(0.31)	−0.06	−0.71	(0.57)	−0.08	−0.13	(0.27)	−0.03
Teaching and interactions	0.78**	(0.29)	0.15	1.48**	(0.55)	0.16	0.52*	(0.26)	0.13
Covariates									
Child age	3.04***	(0.37)	0.49	5.57***	(0.61)	0.51	2.01***	(0.29)	0.42
Child gender (male)	−0.30	(0.24)	−0.06	−0.34	(0.36)	−0.04	0.01	(0.18)	0.00
Time in center	0.06***	(0.02)	0.13	0.11***	(0.03)	0.14	0.03*	(0.01)	0.09
Mother's education	0.24***	(0.05)	0.16	0.41***	(0.08)	0.16	0.14***	(0.04)	0.12
Public center	0.23	(0.41)	0.05	0.31	(0.76)	0.04	0.43	(0.37)	0.11
Urban location	0.88+	(0.45)	0.17	2.13**	(0.83)	0.24	0.90*	(0.40)	0.23
Model 2: interactions: child care quality with parent education									
Provisions × parents' Ed.	−0.76	(0.46)		−0.56	(0.74)		−0.54	(0.38)	
Interactions × parents' Ed.	0.39	(0.47)		0.44	(0.75)		0.38	(0.40)	
Model 3: interactions: child care quality with urban/rural location									
Provisions × urban	0.40	(0.57)		0.85	(1.06)		0.73	(0.49)	
Interactions × urban	−0.38	(0.59)		−1.16	(1.12)		−0.99+	(0.52)	

Notes. (1) Models 2 and 3 include all of variables listed in Model 1 in addition to the interaction terms.
(2) B: regression coefficient estimates; SE: standard error; ES: effect size.
* $p < .05$.
** $p < .01$.
*** $p < .001$.
+ $p < .1$.

Then we used three-level HLMs to analyze the associations between child care quality and child outcomes while controlling for a few important child, family, center, and community characteristics (variable selection was based on results from the preceding analyses) and accounting for the nesting of children in classrooms and classrooms in centers. Results are presented in Table 5. Model 1 tested the main effect model and showed that the children in classrooms with higher CECERS *Teaching and Interactions* scores had significantly higher scores on CDSC language ($d = .15$, $p < .01$), early math ($d = .16$, $p < .01$), and social cognition ($d = .13$, $p < .05$). There was no evidence that the CECERS *Provisions for Learning* score was significantly associated with any child outcome. In addition, children's exposure time in center were also found associated with their developmental outcomes though very modest in effect sizes ($d = .09$–$.14$, $p < .05$).

The second set of analyses determined whether the observed ECE quality on CECERS was differently associated with the outcomes of disadvantaged children who lived in rural areas and had low educated parents. In addition to all of the covariates in Model 1, Model 2 added interactions between the CECERS scores and a dichotomous variable indicating whether both of the child's parents had education levels lower than high school and Model 3 added interactions between the CECERS scores and a dichotomous variable indicating whether the center was located in an urban or rural area. As shown in Table 5, no evidence suggested that parent education level or community type significantly moderated the associations between the CECERS and CDSC scores.

Two sets of subgroup analysis were conducted to verify the robustness of Model 2 and 3's results and to further examine the potential differential effects of ECE quality on subgroups of children. In the first set of subgroup analysis, the study sample was divided into two subgroups by whether both parents' education levels were below high school. In the second set of subgroup analysis, the sample was divided into two subgroups by urban versus rural location. The same Model 1HLMs were conducted for the four subgroups. As shown in Table 6, results from the subgroup analysis indicated that ECE quality of CECERS *Teaching and Interactions* had differential associations with some child outcome variables for certain subgroups of children. The teaching and interaction quality on CECERS had significant and positive associations with early math and social cognition scores for children with less-educated parents and from rural areas, and corresponding effect sizes were

Table 6
Subgroup analyses: testing differential effects of childcare quality on subgroups of children.

	Subgroup	Predictor: CECERS factor score	CDSC language			CDSC early math			CDSC social cognition		
			B	(SE)	ES	B	(SE)	ES	B	(SE)	ES
Subgroup analysis 1: by parent education level	Either ≥ HS	Provisions for learning	−0.03	0.35	−0.01	−0.42	0.64	−0.05	0.10	0.31	0.03
		Teaching and interactions	0.66*	0.33	0.12	1.10+	0.62	0.12	0.33	0.30	0.08
	Both < HS	Provisions for learning	−0.75	0.48	−0.15	−0.73	0.80	−0.08	−0.50	0.40	−0.13
		Teaching and interactions	0.81+	0.46	0.15	1.56*	0.78	0.17	0.81*	0.39	0.20
Subgroup analysis 2: by center location	Urban	Provisions for learning	−0.16	0.50	−0.03	−0.38	0.89	−0.04	0.10	0.40	0.03
		Teaching and interactions	0.56	0.42	0.11	0.92	0.77	0.10	0.21	0.32	0.05
	Rural	Provisions for learning	−0.31	0.42	−0.06	−1.09	0.79	−0.12	−0.51	0.37	−0.13
		Teaching and interactions	0.81+	0.42	0.15	2.09*	0.82	0.23	1.07**	0.38	0.26

Notes. (1) To simplify the table, only the results of the associations between the CECERS factor scores (indicating ECE quality) and CDSC scores (indicating child outcomes) from the final HLM models are presented here. The model covariates include child age, gender, time exposed to ECE, mother's education (in subgroup analysis 2), public child care or otherwise, and urban location or otherwise (in subgroup analysis 1).
(2) In subgroup analysis 1, the two subgroups are defined by children's parent education level. 'Either ≥ HS', at least one parent has high school education or above; 'Both < HS', both parents' education levels are below high school graduation.
* $p < .05$.
** $p < .01$.
+ $p < .1$.

slightly larger than those from whole group analyses. However, this preferable effect for less-advantaged children was not found for language scores. In addition, CECERS factor scores of *Provisions for Learning* had no significant association with any child outcomes in any subgroup, consistent with the results of whole group analyses.

Finally, exploratory analyses were conducted to look for evidence of a threshold effect of ECE quality on child outcomes. We used generalized additive models to empirically explore patterns of association between the CECERS and outcome scores and whether associations between ECE quality and child outcomes differed for the classrooms with different range of CECERS scores. Our analysis yielded no evidence of stronger associations between the CECERS and child outcome scores in specific quality score ranges.

4. Discussion

This study examined ECE quality and its association with child outcomes in one large eastern province in China. As one of the few large-scale studies to examine ECE quality in China, it provided useful information about ECE contexts in the world's most populous country. Our results, consistent with those of much smaller existing studies, further suggest that Chinese ECE is characterized by much larger centers and classrooms and lower quality than is typically described in the U.S. and Northern Europe. Like studies from other countries, we found evidence that Chinese children had higher levels of language, cognitive and social skills when they attended higher quality ECE programs in which teachers engaged in more frequent and positive interactions with children and provided responsive teaching for them. The results of this study did not provide strong and consistent evidence that ECE quality had a differential effect on less-advantaged children based on lower parent education levels and rural community location. We discuss these findings in relation to other studies conducted in Chinese ECE and other contexts as follows.

High rates of kindergarten enrollment in many areas of China have led to questions about the quality of those experiences. The results of this study complement prior reports (Liu, 2011; Pan et al., 2010; Rao et al., 2012; Xiang & Liao, 1995) that Chinese kindergartens have large class sizes and high child-to-adult ratios on average and classrooms are in relatively low quality in terms of access to learning provisions and the process quality of interactions and teaching. These findings probably represent a compromise between cultural beliefs of what is considered good for young children and a focus on increasing access to accommodate the enrollment needs of large numbers of children. Furthermore, the relatively low CECERS scores may reflect a comprehensive effect of (1) traditional beliefs that high quality early education focuses on teaching children rules, maintaining discipline, and helping children to master basic art and academic skills (Li & Rao, 2005; Ng & Rao, 2005; Tobin et al., 2009), (2) high child-to-teacher ratios ($M = 18.23$, SD = 8.59 in this study) and relatively low teacher credentials (38% of all the classroom lead teachers had a BA degree and 77% had teaching certification in this study), and (3) lack of effective ECE facilities and materials that make it difficult for teachers to create opportunities for child-centered learning (as measured by the CECERS *Provisions for Learning* score) and to have responsive teaching and frequent interactions with individual children (as measured by the CECERS *Teaching and Interactions* score).

Although the Chinese Ministry of Education has issued statements about ECE regulations and practices that incorporate Western ECE concepts such as play-based curriculum and child-centered teaching approaches (Ministry of Education of the People's Republic of China, 2001), it is very difficult to implement those policies given current low funding levels and ill-balanced distribution of public funds. According to a recent cost-quality study on 79 kindergartens conducted in the sample province, yearly public funds per child, on average, was roughly USD$354, which took up 45% of the yearly total cost per child. However, the government mainly invested in urban areas and public kindergartens, but significantly less in rural areas and private kindergartens (Li, Pan, & Chen, 2015). Further, low salaries and social statuses (teacher's average monthly salary was roughly USD$398 in this study) make the ECE profession unattractive to potential talents. For large numbers of kindergartens, especially those from rural areas and private sectors, it is difficult to recruit and retain sufficient numbers of qualified teachers who are competent enough to implement developmentally appropriate activities for young children and provide them with quality teaching and interactions in a responsive manner (in this sample, 38% and 35% of classroom lead teachers from private centers and rural areas had no teaching certifications respectively).

Consistent with previous studies' observations (Bu, 2008; Hong & Luo, 2012; Hu, Zhou et al., 2014), ECE quality varies by auspice and type of community/location in this eastern industrial province of China. Our analyses indicate that public kindergartens tend to provide higher quality care and education than private ones, and that rural kindergartens have lower structural and process ECE quality than urban ones. Furthermore, our study also shows that there exist a clear quality gap between kindergartens with overall high parent education and those with low parent education.

More importantly, the results of this study indicate that ECE quality is positively associated with child outcomes in China. The CECERS *Teaching and interactions* scores had a consistently posi-

tive effect on children's language, early math, and social cognition outcomes. Despite the cultural differences in definition of quality childcare and developmentally appropriate activities for young children between China and the Western world, the results of the current study suggest that children show higher levels of language, math/cognitive and social skills when they attend kindergartens of higher quality. Although modest, the effect sizes observed in this study were larger than those reported in recent analyses conducted in the U.S. contexts (Burchinal et al., 2015; Keys et al., 2013).

Previous studies indicated that the quality of teacher–child interactions and direct teaching, compared with environment quality, had more significant impact on child outcomes in language and literacy, pre-academic, and social skills not only in the U.S. contexts (Burchinal et al., 2008) but also in Chinese or related cultures (Li & Rao, 2000). Consistent with these findings, results from the current study suggest that the ECE quality of teaching and interactions, not the provisions for learning, is an effective predictor of children's language, early math, and social learning outcomes. Furthermore, it should be noted that the CECERS factor of *Teaching and Interactions* not only evaluates the quality of interactions between the child and teacher(s) and peers, which is largely equivalent to the scope of *Interaction* subscale of ECERS-R (Harms et al., 1998), but also addresses the process quality of whole-group teaching in Chinese kindergarten classrooms. Rated based on direct observation of the first group-teaching activity during the school day, the *Whole-Group Instruction* subscale of CECERS consists of 7 items, which define what is considered to be high quality group teaching in Chinese ECE contexts: appropriate teaching objectives and deliberately selected teaching content with considering for developmental differences between individuals; well designed and organized teaching process in accordance with children's interest and learning approaches; timely and positive responses to students' emotional and learning needs; providing sufficient support, necessary scaffolding, and frequent feedback for children's learning, especially for their language and thinking skill development; children's engagement and performance during the teaching activity (Li & Hu, 2012). On the whole, high quality whole-group teaching can be regarded as a constructive learning process for a group of children to know, to experience, and to explore with necessary support from teacher(s) and frequent interactions with teacher(s) and peers (Li et al., 2014). The findings from the current study suggested that high quality whole-group teaching might play a significant role in the development of Chinese preschoolers, especially for their learning outcomes in language, cognitive, and social skills.

Time in formal childcare (i.e., center care) has shown positive associations with child outcomes, especially in (pre-) academic skills, in other studies across different countries (Burchinal, Peisner-Feinberg, Pianta, & Howes, 2002; Pinto, Pessanha, & Aguiar, 2013; Rao et al., 2012). In line with these findings, this study also suggested that the exposure time to ECE experience in kindergartens promotes early development of Chinese children as the variable of 'time in center' was found significantly associated with all the three child outcomes though very modest in effect size.

Many U.S. studies have suggested that low-income and at-risk children may benefit more from quality ECE (Burchinal, Peisner-Feinberg, Bryant, & Clifford, 2000; Gormley, Gayer, Phillips, & Dawson, 2005; Votruba-Drzal, Coley, & Chase-Lansdale, 2004; Winsler et al., 2008). However, this study did not provided strong and consistent evidence to support these findings. The interaction models did not find that ECE quality was significantly moderated by parent education level or type of community/location, but subgroup analysis indicated that ECE quality of teaching and interactions could be a predictor of cognitive and social developmental outcomes for rural but not urban children and for children with lower but not higher educated parents. Although we favor the results of interaction models because they used a larger sample, provided direct statistical tests for differences between subgroups, and the subgroup analysis results could just be an artifact of this specific sample (Assmann, Pocok, Enos & Kasten, 2000), it is possible that our sample was underpowered to detect the moderation effects of interest. In many empirical studies, the magnitude of an interaction is smaller than that of a main effect. This is the case for our study. A larger, sometimes significantly larger, sample is usually required to detect interaction effects in such a case (Brookes et al., 2001). The question of differential ECE effects on subpopulations is important and warrants further investigations. Larger studies currently being conducted in China will have greater power to test these interactions.

No evidence emerged to indicate a threshold effect in the associations between ECE quality and child outcomes in this study. Evidence of thresholds have been detected in other studies (Burchinal, Vandergrift, Pianta, & Mashburn, 2010; Burchinal, Vernon-Feagans, Vitiello, Greenberg, & The Family Life Project Key Investigators, 2014; Leyva et al., 2015), albeit somewhat inconsistently. However, those studies tended to have larger sample of ECE programs with higher quality and/or measured quality based on other measures (Burchinal et al., 2015).

4.1. Implications for policymakers and practitioners

The findings of this study suggest that ECE quality is positively associated with child outcomes in China. We obtained limited evidence that ECE quality might matter more for disadvantaged Chinese children and this issue must be further investigated in future studies. The current study conveys valuable messages to Chinese policymakers who are seeking empirical evidence to appropriately allocate public ECE investments and other resources to support child development. Investment should be directed toward improving teacher credentials and practices that improve process quality, such as increasing ECE teacher certification standards, building up effective technical assistance system in support of ECE teachers' professional development, and including process quality indicators in the governmental ECE program rating system. Our research also calls for more rigorous studies that examine differential effects of ECE on subpopulations and these results can be used to inform ECE policies that benefit all Chinese children.

As indicated in many studies conducted in the U.S. (Hamre et al., 2012; Hamre, Hatfield, Pianta, & Jamil, 2014) and in China (Kang, Liu, & Liu, 2011), teachers play a significant role in providing high quality teaching and interactions. In fact, the biggest and most consistent challenges for rural kindergartens are to recruit and retain quality ECE teachers (Hu & Roberts, 2013). More qualified teachers are needed to teach in rural kindergartens, improve program quality, and narrow the quality gap between urban and rural areas. The Chinese government must take key steps to improve the ECE system in rural settings and ensure a more balanced supply of quality teachers across the country, such as providing tenure status and subsidies for qualified ECE teachers who work in rural areas and in private sectors and setting up effective teacher-exchange systems between urban and rural areas. Meanwhile, it is important for kindergartens, especially private ones, to recruit and retain qualified teachers and provide effective professional development to improve classroom practices, such as how to implement developmentally appropriate high quality group teaching and other learning activities, support children's free play and outdoor activities, and interact with children in a more frequent, responsive, and individualized manner.

4.2. Limitations

The current study had several limitations. First, this was an observational study, and we could not make causal conclusions

about the association between CECERS scores and child outcomes. We did not measure children's baseline developmental scores, so we could not include them as an important control variable. Only a few family demographic variables were collected from most families, also limiting our ability to control for baseline differences among the children prior to entering kindergartens. It is likely that parents with more money and more education provide many opportunities, including higher quality ECE, to their children that result in better outcomes, and the observed associations in this study are likely inflated because they may reflect those other unmeasured opportunities. Ideally, econometric methods such as propensity score matching could have reduced the impact of such selection bias in this study, but the lack of baseline information on the child and limited information about the family meant that those approaches would not provide much correction for potential selection bias.

Second, concerns about the child assessment tool also limit the conclusions from this study. The child outcome measure was 20 years old at the time of data collection in the current study. Although researchers reported good psychometric properties when the measure was initially published, it was unclear whether a similar level of good psychometric properties would still apply to a 2012 sample. However, the CDSC was the only available child developmental measure developed for and validated in the context of China at the time of this study. We believe that the scores were more objective than teacher ratings because our testers went through extensive training on administering CDSC and all items in CDSC are very clearly defined. There was little reason to believe that the measure was biased due to the use of trained independent testers.

Third, although the sample was large and diverse, it was not a technically population-representative sample. We acknowledge that the findings may not be sufficiently representative of the province in which the study was conducted. However, our sample included all of the major demographic strata of kindergartens in the province, and it was much larger and more diverse than most previous studies conducted in China.

In addition, age in HLMs of this study is used as a continuous, not a categorical variable. There might be interaction or differential effects between different age groups and ECE quality and child outcome measures. This might be a potential limitation of the current study and need to be further examined.

4.3. Summary

This study aimed to examine ECE quality and its associations with child developmental outcomes in an eastern province of China. ECE quality was higher in urban than rural areas, in public than private kindergartens, and in programs with high parent education than those with low parent education, although quality was relatively low overall. One dimension of quality, teaching and interactions, was consistently associated with child outcomes in language, early math, and social cognition. The study contributes to the growing body of international literature on associations between ECE quality and child outcomes in China. It also raises additional questions to be further explored.

Acknowledgement

This study is supported by the Grant: "A Multidimensional and Integrated View Approach to Chinese Early Care and Education Quality Evaluation" from the National Social Science Foundation of China (Kejian Li; Grant #: CHA110131).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ecresq.2016.01.009.

References

Aboud, F. E. (2006). Evaluation of an early childhood preschool program in rural Bangladesh. *Early Childhood Research Quarterly*, *21*(1), 46–60. http://dx.doi.org/10.1016/j.ecresq.2006.01.008

Aboud, F. E., & Hossain, K. (2011). The impact of preprimary school on primary school achievement in Bangladesh. *Early Childhood Research Quarterly*, *26*(2), 237–246. http://dx.doi.org/10.1016/j.ecresq.2010.07.001

Assmann, S. F., Pocock, S. J., Enos, L. E., & Kasten, L. E. (2000). Subgroup analysis and other (mis) uses of baseline data in clinical trials. *The Lancet*, *355*(9209), 1064–1069. http://dx.doi.org/10.1016/S0140-6736(00) 02039-0

Auger, A., Farkas, G., Burchinal, M. R., Duncan, G. J., & Vandell, D. L. (2014). Preschool center care quality effects on academic achievement: An instrumental variables analysis. *Developmental Psychology*, *50*(12), 2559–2571. http://dx.doi.org/10.1037/a0037995

Baker-Henningham, H., Walker, S., Powell, C., & Gardner, J. M. (2009). A pilot study of the Incredible Years Teacher Training programme and a curriculum unit on social and emotional skills in community pre-schools in Jamaica. *Child: Care, Health & Development*, *35*(5), 624–631. http://dx.doi.org/10.1111/j.13652214.2009.00964.x

Bronfenbrenner, U. (1979). *The ecology of human development: experiments in nature and design*. Cambridge, MA: Harvard University Press.

Bronfenbrenner, U., & Morris, P. A. (1998). The ecology of developmental processes. In W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology, Vol. 1: theoretical models of human development* (5th ed., pp. 993–1023). New York, NY: Wiley.

Brookes, S. T., Whitley, E., Peters, T. J., Mulheran, P. A., Egger, M., & Davey Smith, G. (2001). Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and false-negatives. *Health Technology Assessment*, *5*(33), 1–56. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11701102

Bryant, D. M., Zaslow, M., & Burchinal, M. (2010). Issues in measuring program quality. In P. W. Wesley, & V. Buysse (Eds.), *The quest for quality: promising innovations for early childhood programs*. Baltimore MD: Paul H. Brookes.

Bu, S. M. (2008). The agony caused by preschool education marketization: take some areas of Zhejiang Province as examples. *Educational Development*, *20*, 22–26. Chinese language.

Burchinal, M., Howes, C., Pianta, R., Bryant, D., Early, D., Clifford, R., & Barbarin, O. (2008). Predicting child outcomes at the end of kindergarten from the quality of pre-kindergarten teacher–child interactions and instruction. *Applied Developmental Science*, *12*(3), 140–153. http://dx.doi.org/10.1080/10888690802199418

Burchinal, M., Kainz, K., & Cai, Y. (2011). How well do our measures of quality predict child outcomes? In M. Zaslow, I. Martinez-Beck, K. Tout, & T. Halle (Eds.), *Quality measurement in early childhood settings* (pp. 11–31). Baltimore, MD: Paul H. Brookes.

Burchinal, M., Magnuson, K., Powell, D., & Hong, S. L. (2015). Early child care and education. In *Handbook of child psychology*. New York, NY: Wiley.

Burchinal, M. R., Peisner-Feinberg, E., Bryant, D. M., & Clifford, R. (2000). Children's social and cognitive development and child-care quality: testing for differential associations related to poverty, gender, or ethnicity. *Applied Developmental Science*, *4*(3), 149–165. http://dx.doi.org/10.1207/S1532480XADS0403_4

Burchinal, M. R., Peisner-Feinberg, E., Pianta, R., & Howes, C. (2002). Development of academic skills from preschool through second grade: family and classroom predictors of developmental trajectories. *Journal of School Psychology*, *40*(5), 415–436. http://dx.doi.org/10.1016/S0022-4405(02) 00107-3

Burchinal, M., Vandergrift, N., Pianta, R., & Mashburn, A. (2010). Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs. *Early Childhood Research Quarterly*, *25*(2), 166–176. http://dx.doi.org/10.1016/j.ecresq.2009.10.004

Burchinal, M., Vernon-Feagans, L., Vitiello, V., Greenberg, M., & The Family Life Project Key Investigators. (2014). Thresholds in the association between child care quality and child outcomes in rural preschool children. *Early Childhood Research Quarterly*, *29*(1), 41–51. http://dx.doi.org/10.1016/j.ecresq.2013.09.004

Chen, D., Hu, B., Fan, X., & Li, K. (2014). Measurement quality of the *Chinese Early Childhood Program Rating Scale (CECPRS)*: an investigation using multivariate generalizability theory. *Journal of Psychoeducational Assessment*, *32*(3), 236–248. http://dx.doi.org/10.1177/0734282913504813

Engle, P. L., Fernald, L. C., Alderman, H., Behrman, J., O'Gara, C., Yousafzai, A., & Iltus, S. (2011). Strategies for reducing inequalities and improving developmental outcomes for young children in low-income and middle-income countries. *The Lancet*, *378*(9799), 1339–1353. http://dx.doi.org/10.1016/S0140-6736(11) 60889-1

Goldstein, H. (2009). *REALCOM-IMPUTE: multiple imputation using MLwiN*. Center for Multilevel Modelling, University of Bristol.

Gordon, R. A., Fujimoto, K., Kaestner, R., Korenman, S., & Abner, K. (2013). An assessment of the validity of the ECERS-R with implications for measures of

child care quality and relations to child development. *Developmental Psychology*, *49*(1), 146–160. http://dx.doi.org/10.1037/a0027899
Gormley, W. T., Gayer, T., Phillips, D., & Dawson, B. (2005). The effects of universal pre-k on cognitive development. *Developmental Psychology*, *41*(6), 872–884. http://dx.doi.org/10.1037/0012-1649.41.6.872
Guo, Y. H. (2008). On the causes of rural-urban disparities and the options of developmental target for rural-urban integration. *Forum of Southern China*, *3*, 58–65. Chinese language.
Hamre, B., Hatfield, B., Pianta, R., & Jamil, F. (2014). Evidence for general and domain-specific elements of teacher–child interactions: associations with preschool children's development. *Child Development*, *85*(3), 1257–1274. http://dx.doi.org/10.1111/cdev.12184
Hamre, B. K., Pianta, R. C., Burchinal, M., Field, S., LoCasale-Crouch, J., Downer, J. T., . . . & Scott-Little, C. (2012). A course on effective teacher–child interactions effects on teacher beliefs, knowledge, and observed practice. *American Educational Research Journal*, *49*(1), 88–123. http://dx.doi.org/10.3102/0002831211434596
Harms, T., Clifford, R. M., & Cryer, D. (1998). *Early Childhood Environment Rating Scale—revised edition (ECERS-R)*. New York, NY: Teachers College Press.
Heckman, J. J. (2010). Building bridges between structural and program evaluation approaches to evaluating policy. *Journal of Economic Literature*, *48*(2), 356–398. http://dx.doi.org/10.1257/jd.48.2.356
Hong, X. M., & Luo, L. (2012). Analyzing the rural-urban disparities in Chinese preschool education from the notion of equity. *Journal of Pedagogy*, *10*, 73–81. Chinese language.
Hu, B. Y. (2014). Comparing cultural differences in two quality measures in Chinese kindergartens: The Early Childhood Environment Rating Scale-revised and the kindergarten quality rating system. *Compare: A Journal of Comparative and International Education*, *45*(1), 94–117. http://dx.doi.org/10.1080/03057925.2013.841468
Hu, B., & Li, K. (2012). The quality rating system of Chinese preschool education: prospects and challenges. *Childhood Education*, *88*(1), 14–22. http://dx.doi.org/10.1080/00094056.2012.643711
Hu, B., & Roberts, S. K. (2013). The transformation of Chinese rural preschools: retrospect and prospect. *International Journal of Educational Development*, *33*(4), 316–324. http://dx.doi.org/10.1016/j.ijedudev.2012.04.006
Hu, B. Y., Vong, K., Chen, Y., & Li, K. (2014). Expert practitioner's views about the Chinese early childhood environment rating scale. *European Early Childhood Education Research Journal*, *23*(2), 229–249. http://dx.doi.org/10.1080/1350293X.2014.919779
Hu, B. Y., Zhou, Y., Li, K., & Roberts, S. K. (2014). Examining program quality disparities between urban and rural kindergartens in China: evidence from Zhejiang. *Journal of Research in Childhood Education*, *28*(4), 461–483. http://dx.doi.org/10.1080/02568543.2014.944720
Kang, J. Q., Liu, Y., & Liu, P. (2011). An analysis of the investment efficiency of the preschool year education in rural area. *Educational Research Monthly*, *5*, 34–37. Chinese language.
Kenward, M. J., & Carpenter, J. R. (2008). Multiple imputation. In G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Molenberghs (Eds.), *Longitudinal data analysis: a handbook of modern statistical methods* (pp. 477–499). Boca Raton, FL: Chapman & Hall/CRC.
Keys, T. D., Farkas, G., Burchinal, M. R., Duncan, G. J., Vandell, D. L., Li, W., & Howes, C. (2013). Preschool center quality and school readiness: quality effects and variation by demographic and child characteristics. *Child Development*, *84*(4), 1171–1190. http://dx.doi.org/10.1111/cdev.12048
Leyva, D., Weiland, C., Barata, M., Yoshikawa, H., Snow, C., Treviño, E., & Rolla, A. (2015). Teacher–child interactions in Chile and their associations with prekindergarten outcomes. *Child Development*, *86*(3), 781–799. http://dx.doi.org/10.1111/cdev.12342
Li, H., & Rao, N. (2000). Parental influences on Chinese literacy development: a comparison of preschoolers in Beijing, Hong Kong, and Singapore. *International Journal of Behavioral Development*, *24*(1), 82–90.
Li, H., & Rao, N. (2005). Influences on literacy attainment: evidence from Beijing, Hong Kong and Singapore. *International Journal of Early Years Education*, *13*(3), 235–253. http://dx.doi.org/10.1080/09669760500295870
Li, K., Hu, B., Pan, Y., Qin, J., & Fan, X. (2014). Chinese Early Childhood Environment Rating Scale (trial) (CECERS): a validity study. *Early Childhood Research Quarterly*, *29*(3), 268–282. http://dx.doi.org/10.1016/j.ecresq.2014.02.007
Li, K. J., & Hu, B.Y. (2012). Chinese Early Childhood Environment Rating Scale (CECERS; trial version), Unpublished instrument under development (Chinese language).
Li, K. J., Pan, Y., & Chen, Q. X. (2015). A study on the correlations between quality and annual average investment/cost per student of early childhood education. *Education & Economy*, *2*, 25–31. Chinese language.
Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). New York: Wiley.
Liu, Z. L. (2011). *Evaluating kindergarten quality in China: an investigation in 11 provinces*. Beijing, China: Educational Science Publishing House. Chinese language.
Malmberg, L. E., Mwaura, P., & Sylva, K. (2011). Effects of a preschool intervention on cognitive development among East-African preschool children: a flexibly time-coded growth model. *Early Childhood Research Quarterly*, *26*(1), 124–133. http://dx.doi.org/10.1016/j.ecresq.2010.04.003
Ministry of Education of the People's Republic of China (2001). Kindergarten education guideline (trial). Retrieved from http://www.moe.gov.cn/publicfiles/business/htmlfiles/moe/s3327/201001/81984.html (Chinese language).
Ministry of Education of the People's Republic of China (2013). Number of kindergartens, classes in preschool education. Retrieved from http://www.moe.gov.cn/publicfiles/business/htmlfiles/moe/s7567/201308/156409.html (Chinese language)
Moore, A. C., Akhter, S., & Aboud, F. E. (2008). Evaluating an improved quality preschool program in rural Bangladesh. *International Journal of Educational Development*, *28*(2), 128–131. http://dx.doi.org/10.1016/j.ijedudev.2007.05.003
National Bureau of Statistics of the People's Republic of China. (2013). *2012 China Statistical Yearbook*. Beijing, China: Publisher of National Bureau of Statistics of the People's Republic China. Chinese language.
Ng, S. S. N., & Rao, N. (2005). Teaching mathematics in Hong Kong: a comparison between the pre-primary and early primary years. *Hong Kong Journal of Early Childhood*, *4*(1), 30–36. Retrieved from http://repository.ied.edu.hk/dspace/handle/2260.2/3545
Pan, Y., Liu, Y., & Lau, E. (2010). Evaluation of the kindergarten quality rating system in Beijing. *Early Education and Development*, *2*(2), 186–204. http://dx.doi.org/10.1080/10409281003631134
Pang, Z. Q. (2006). An empirical study on rural-urban disparities within Gansu Province. *Social Science Review*, *4*, 15–19. Chinese language.
Peisner-Feinberg, E. S., & Yazejian, N. (2010). Research on program quality: the evidence base. In V. Buysse, & P. W. Wesley (Eds.), *The quest for quality: promising innovations for early childhood programs* (pp. 21–45). Baltimore, MD: Paul H. Brookes.
Pinto, A. I., Pessanha, M., & Aguiar, C. (2013). Effects of home environment and center-based child care quality on children's language, communication, and literacy outcomes. *Early Childhood Research Quarterly*, *28*(1), 94–101. http://dx.doi.org/10.1016/j.ecresq.2012.07.001
Rao, N. N. (2010). Preschool quality and the development of children from economically disadvantaged families in India. *Early Education and Development*, *21*(2), 167–185. http://dx.doi.org/10.1080/10409281003635770
Rao, N., Sun, J., Zhou, J., & Zhang, L. (2012). Early achievement in rural China: The role of preschool experience. *Early Childhood Research Quarterly*, *27*(1), 66–76. http://dx.doi.org/10.1016/j.ecresq.2011.07.001
Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: John Wiley & Sons, Inc.
San Francisco, A. R., Arias, M., Villers, R., & Snow, C. (2006). Evaluating the impact of early literacy interventions on low-income Costa Rican kindergarteners. *International Journal of Educational Research*, *45*(3), 188–201. http://dx.doi.org/10.1016/j.ijer.2006.11.002
SAS Institute. (2011). *SAS 9.3*. Cary, NC: SAS Institute.
StataCorp. (2013). *Stata Statistical Software: Release 13*. College Station, TX: StataCorp LP.
Stipek, D., & Byler, P. (2004). The early childhood classroom observation measure. *Early Childhood Research Quarterly*, *19*, 375–397. http://dx.doi.org/10.1016/j.ecresq.2004.07.007
Sylva, K., Siraj-Blatchford, I., & Taggart, B. (2003). *Assessing quality in the early years: Early Childhood Environment Rating Scale-Extension (ECERS-E)*. Stoke-on-Trent, England: Trentham Books.
Tang, T., & Zhao, L. Q. (2008). An empirical study on rural-urban gap in social economic development in Chongqing Province. *China Economist*, *7*, 276–277. Chinese language.
The Central Government of the People's Republic of China (2010). Compendium for China's mid-and long-term education reform and development: 2010–2020. Retrieved from http://www.gov.cn/jrzg/2010-07/29/content_1667143.htm (Chinese language).
The State Council of the People's Republic of China (2010). Chinese language. The State Council's several suggestions regarding developing preschool education. Retrieved from http://www.moe.edu.cn/publicfiles/business/htmlfiles/moe/s3327/201011/111850.html.
Tobin, J., Hsueh, Y., & Karasawa, M. (2009). *Preschool in three cultures revisited: China, Japan and the United States*. Chicago, IL: University of Chicago Press.
Votruba-Drzal, E., Coley, R. L., & Chase-Lansdale, P. L. (2004). Child care and low-income children's development: direct and moderated effects. *Child Development*, *75*(1), 296–312. http://dx.doi.org/10.1111/j.1467-8624.2004.00670.x
Winsler, A., Tran, H., Hartman, S. C., Madigan, A. L., Manfra, L., & Bleiker, C. (2008). School readiness gains made by ethnically diverse children in poverty attending center-based childcare and public school pre-kindergarten programs. *Early Childhood Research Quarterly*, *23*(3), 314–329. http://dx.doi.org/10.1016/j.ecresq.2008.02.003
Wu, K. B., Young, M. E., & Cai, J. (2012). *Early child development in China: breaking the cycle of poverty and improving future competitiveness*. Washington, DC: World Bank., retrieved from: https://openknowledge.worldbank.org/handle/10986/9383
Xiang, Z. P. (1995). On the issues and evaluation orientation of early childhood education process—based on assessment of early childhood education programs in six provinces of China. *Studies in Preschool Education*, *2*, 31–35. Chinese language.
Xiang, Z. P., & Liao, Y. (1995). *An evaluation study on early childhood education program quality in six provinces of China*. Beijing,China: Educational: Science Publishing House. Chinese language.
Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinosa, L. M., Gormley, W. T., . . . & Zaslow, M. J. (2013). Investing in our future: the evidence base on preschool education. *Society for Research in Child Development and Foundation for Child Development*.

Zaslow, M., Martinez-Beck, I., Tout, K., & Halle, T. (2011). *Quality measurement in early childhood settings*. Baltimore, MD: Paul H. Brookes.
Zhang, H. C., Zhou, R., Chen, G. M., Zhao, Z. M., & Wang, X. P. (1992). *Child Developmental Scale of China (age 3 to 6)*. Beijing, China: Child Development Center of China. Chinese language.
Zhang, Y. Y., Luo, F., Tao, S., Luo, L., & Dong, Q. (2015). Family socioeconomic status, parental investment and migrant children's academic achievement in China. *Journal of Psychological Science*, *38*(1), 19–26. Chinese language.
Zhou, R., & Zhang, H. C. (1994). The development of Child Developmental Scale of China (age 3 to 6). *The Science of Psychology*, *17*(3), 137–192. Chinese language.
Zhu, J., & Wang, X. C. (2005). Contemporary early childhood education and research in China. In B. Spodek, & O. N. Saracho (Eds.), *International perspectives on research in early childhood education* (pp. 55–77). Greenwich, CT: Information Age Pub.
Zhu, J., & Zhang, J. (2008). Contemporary trends and developments in early childhood education in China. *International Journal of Early Years Education*, *28*(20), 173–182. http://dx.doi.org/10.1080/09575140802163584