Early Childhood Research Quarterly 29 (2014) 268–282

Contents lists available at ScienceDirect

Early Childhood Research Quarterly

Chinese Early Childhood Environment Rating Scale (*trial*) (CECERS): A validity study

Kejian Li[a,1], Bi Ying Hu[c,1], Yi Pan[b], Jinliang Qin[a], Xitao Fan[c,*]

[a] *Zhejiang Normal University, China*
[b] *University of North Carolina at Chapel Hill, USA*
[c] *University of Macau, Macao*

ARTICLE INFO

Article history:
Received 17 March 2013
Received in revised form 19 February 2014
Accepted 24 February 2014
Available online 21 March 2014

Keywords:
Early childhood education
Quality measurement
Reliability
Validity

ABSTRACT

The *Chinese Early Childhood Environment Rating Scale (trial) (CECERS)* is a new instrument for measuring early childhood program quality in the Chinese socio-cultural contexts, based on substantial adaptation from the *Early Childhood Environment Rating Scale-Revised Edition (ECERS-R)*. This paper describes the development and validation process of *CECERS*. Empirical data were collected from a stratified random sample 178 classrooms, from which a random sample of 1012 children was measured for child development outcomes. Guided by the framework of broad conceptualization of validity and validation as advocated by Messick (1989), evidence in a variety of forms is presented and discussed, including content validity considerations (e.g., measuring socially and culturally relevant domains), measurement reliability considerations (e.g., internal consistency reliability, inter-rater reliability), and measurement validity considerations (concurrent validity, criterion-related validity, internal structure based on exploratory factor analysis). The empirical findings for *CECERS* compare very favorably with the validation outcomes of *ECERS-R*. The body of evidence accumulated in the validation process supports the use and interpretation of *CECERS* scores as quality indicators of early childhood education program in the Chinese social and cultural contexts. Limitations and future directions are also discussed.

1. Introduction

In China, kindergartens (or *you'eryuan* in Chinese) represent the most predominant early childhood education (ECE) programs serving children of ages 3–6. A kindergarten is a school with multiple classrooms for kids, with kids typically being at about the same age in the same classroom. Typically, a child enters the kindergarten at the age of 3 and spends three years with the same group of peers before he/she goes to primary school. Over the past two decades, significant progress has been made in China in ECE services, with about 62.3% of 48 million children (age 3–6) served in 166,750 registered kindergartens in 2012, relative to about 27.8% in 1991 (National Bureau of Statistics of the People's Republic of China, 2012). Historically, most of the public funding in ECE has been in more affluent urban areas (Wu, Young, & Cai, 2012), resulting in a staggeringly wide variation in ECE program quality across regions and across urban and rural programs. For example, a recent study revealed that student-teacher ratio has an average of about 10:1 in Shanghai and Beijing, the most developed metropolitan areas in China, but 94:1 in some rural areas of Guizhou, one of the poorest provinces in southwest China (Wu et al., 2012). Teacher qualifications also vary widely across cities, towns, and rural areas (Hu & Roberts, 2013; Wu et al., 2012).

The Central Government of the People's Republic of China (2010) has committed to achieving the goal of education equity in ECE, by using financial and policy leverages (e.g., subsidies for public kindergartens and teacher training in rural area; Hu & Li, 2012) in favor of socially disadvantaged children. For achieving these national goals, sound ECE program quality measures are needed to help make financial and policy decisions. As a result, the development of a culturally appropriate and psychometrically sound measure of ECE program quality has become a task of high importance. Since the late 1980s, local education authorities (LEAs) have established Kindergarten Quality Rating Systems (KQRS), primarily based on the evaluation criteria from ECE policies and regulations issued by the National Education Committee (NEC, 1990, 1996) and the Ministry of Education (MOE, 2001). However, these KQRS generally emphasize structural quality, such as management aspects of

* Corresponding author at: Faculty of Education, University of Macau, Macao. Tel.: +853 8397 8773; fax: +853 2883 1695.
E-mail address: xtfan@umac.mo (X. Fan).
[1] The first two authors have equal contributions to this article.

http://dx.doi.org/10.1016/j.ecresq.2014.02.007
0885-2006/

program quality, at the expense of process quality, such as planning the learning environment and curricular activities and student-teacher interactions (Dai & Liu, 2003; Hu & Li, 2012; Liu, 1998; Pan, Liu, & Lau, 2010). In addition, these KQRS have unknown psychometric quality, as they have not gone through any psychometric validation processes (Li & Hu, 2012).

In fact, in China that has the largest population of children in need of ECE, there is a severe paucity of research regarding ECE quality; as a result, there is lack of research base to guide policy formation and classroom practices (Hu & Li, 2012). In response to this, Liu and Pan developed the *Kindergarten Environment Rating Scale* (KERS; Liu & Pan, unpublished), inspired by the *Early Childhood Environment Rating Scale-Revised Edition* (ECERS-R; Harms, Clifford, & Cryer, 2005) and the *Early Childhood Environment Rating Scale-Extension* (ECERS-E; Sylva, Siraj-Blatchford, & Taggart, 2003). Liu and Pan (2008) reported acceptable evidence of psychometric quality of the KERS, which assesses 19 items organized in four subscales (Liu & Pan, 2008). Li and Xu (2012), however, identified several major shortcomings of the KERS, such as incomplete representation of ECE quality concepts and inadequate measurement of process quality. Moreover, KERS remains an unpublished measure, and has not been adopted by professionals of ECE in China.

The State Council of the People's Republic of China (2010) called upon the ECE professional community to develop a socially, culturally, and psychometrically appropriate measure of Chinese ECE program quality. In this article, we describe the development of the *Chinese Early Childhood Environment Rating Scale (trial)* (*CECERS*), an ECE program quality measure adapted from the *ECERS-R* and designed for the cultural contexts in China, and we present the empirical evidence for the psychometric properties of *CECERS*.

1.1. ECERS-R in international contexts

In the U.S., the widely used *ECERS-R* (Harms et al., 2005) is the only instrument available for measuring the overall or global quality of an early childhood program, even though there are a number of tools for measuring some specific aspects of ECE program quality. The original *ECERS* (Harms & Clifford, 1980) and its revised versions have been used in numerous large-scale studies about ECE program quality (Early et al., 2006; Lambert, Abbott-Shim, Sibley, Spodek, & Saracho, 2006; Phillips, Mekos, Scarr, McCartney, & Abbott-Shim, 2000). The *ECERS-R* has been adopted by nearly all of the current Quality Rating and Improvement Systems (QRIS) in the U.S. (Tout et al., 2010). Research findings related to the use of *ECERS-R* have generated far-reaching impact on ECE practice, research, and policy development (Fenech, 2011).

By far, the *ECERS-R* is the most well researched and widely used instrument for measuring ECE quality in international contexts (Gol-Guven, 2009). So far, different versions of *ECERS-R* have been applied in over 20 countries of different socioeconomic backgrounds (Myers, 2004). However, one main difficulty in using the measure in cross-cultural contexts is that a measure based on the standards and expectations relevant to one culture may not be appropriate for another social and cultural context (Limlingan, 2011). A host of issues has been raised by different researchers in using *ECERS* or *ECERS-R* in different social and cultural contexts (Lambert et al., 2008; Mathers, Linskey, Seddon, & Sylva, 2007; Sakai, Whitebook, Wishard, & Howes, 2003; Sheridan, Giota, Han, & Kwon, 2009). In England, the concern that *ECERS* did not meet the evaluation needs in that social and cultural context led to the development of *ECERS-E* (Sylva et al., 2006).

Furthermore, various changes and adaptations have been made while using *ECERS*/*ECERS-R* in different cultural/social contexts, such as elimination of selected constructs for Chilean (Villalon, Suzuki, Herrera, & Mathiesen, 2002) and Korean (Lim, 1983) contexts, reduction of the number of items for use in Sweden (Kärrby, 1989; Kärrby & Giota, 1994), removal of socially/culturally irrelevant items and addition of new and culturally relevant items for Bangladesh environment (Aboud, 2006), and much substantial modifications for use in India (Isley, 2000). From these studies involving the use of *ECERS*/*ECERS-R* in other social/cultural contexts, there is the general consensus that changes and modifications are necessary for enhancing the relevance and appropriateness of the measure for the different social and cultural contexts.

Although originally designed for the U.S. contexts, the *ECERS*/*ECERS-R* showed good psychometric properties in terms of evidence for measurement reliability and validity in both U.S. and some other socio-cultural contexts (Harms et al., 2005; Mathers et al., 2007; Rentzou, 2010; Tietze, Cryer, Bairrão, Palacios, & Wetzel, 1996). However, research studies about the underlying factor structure of the *ECERS*/*ECERS-R* yielded inconsistent findings, ranging from one single-dimension global quality factor (Holloway, Kagan, Fuller, Tsou, & Carroll, 2001; Scarr, Eisenberg, & Deater-Deckard, 1994), a two-factor structure (Cassidy, Hestenes, Hegde, Hestenes, & Mims, 2005; Sakai et al., 2003), to a three-factor structure (Gordon, Fujimoto, Kaestner, Korenman, & Abner, 2013; Perlman, Zellman, & Le, 2004). Even for studies suggesting the same number of factors (e.g., a two-factor structure), the composition of the factors could be quite different (Cassidy et al., 2005; Sakai et al., 2003). The empirical inconsistencies about the underlying factor structure of the *ECERS*/*ECERS-R* indicate that it could be practically very challenging in conceptualizing the "latent" statistical constructs for a measure of childhood education program quality like *ECERS*/*ECERS-R*. In addition, Gordon et al. (2013) also identified and discussed issues of "losing of quality information" and "rating category disordering" as a result of the "stop scoring procedure" used in the *ECERS-R*. Gordon et al. (2013) further suggested scoring all indicators and treating the *ECERS-R* as a checklist rather than a scale.

The current study can be considered as the continuation of efforts in adapting the *ECERS*/*ECERS-R* for different social and cultural contexts on the international stage. As described below, based on a rigorous pilot study of the *ECERS-R*, and on the re-conceptualization about program quality of ECE in the social and cultural contexts of China, the *Chinese Early Childhood Environment Rating Scale (trial)* (*CECERS*) was developed based on substantial adaptation of the *ECERS-R*. In the following, we elaborate on the development process of the *CECERS*, and its similarities and differences compared with the *ECERS-R*.

1.2. The development of CECERS

1.2.1. A pilot study of applying ECERS-R in Chinese kindergartens

To understand how well the concepts of childhood education program quality represented in *ECERS-R* were congruent with the Chinese societal and cultural contexts, in 2010, *ECERS-R* was used to measure the program quality in 105 kindergarten classrooms in China (Li, Hu, Pan, & Qin, 2013). The findings revealed that some quality concepts for childhood education program (e.g., space for privacy, substantial amount of time for free play, acceptance of diversity) were not endorsed by Chinese childhood education practitioners. On the other hand, some quality concepts valued in the Chinese culture (e.g., those related to collectivism and collective activities) were not represented in the *ECERS-R*. Psychometric analysis also revealed problems related to multiple subscales of *ECERS-R* (e.g., *Activities*, *Language-Reasoning*, *Program Structure*, and *Interaction*). These findings, together with the feedback from the practitioners and raters, suggested that, without any adaptation, *ECERS-R* would lack cultural appropriateness as a tool for measuring the quality of Chinese ECE programs (Li et al., 2013). There were multiple reasons for the lack of cultural appropriateness of *ECERS-R* in the Chinese contexts, but a major reason was that, the *ECERS-R*