Research Paper

Using the early childhood environment rating scale-Revised in high stakes contexts: Does evidence warrant the practice?

Claude Messan Setodji[a,*], Diana Schaack[b], Vi-Nhuan Le[c]

[a] RAND Corporation, 4570 5th Avenue, Suite 600, Pittsburgh, PA 15213, United States
[b] University of Colorado Denver, United States
[c] NORC at the University of Chicago, United States

ARTICLE INFO

ABSTRACT

Increasingly, states establish different thresholds on the Early Childhood Environment Rating Scale–Revised (ECERS–R), and use these thresholds to inform high-stakes decisions. However, the validity of the ECERS-R for these purposes is not well established. The objective of this study is to identify thresholds on the ECERS-R that are associated with preschool-aged children's social and cognitive development. Applying non-parametric modeling to the nationally-representative Early Childhood Longitudinal Study Birth Cohort (ECLS-B) dataset, we found that once classrooms achieved a score of 3.4 on the overall ECERS-R composite score, there was a leveling-off effect, such that no additional improvements to children's social, cognitive, or language outcomes were observed. Additional analyses found that ECERS-R subscales that focused on teaching and caregiving processes, as opposed to the physical environment, did not show leveling-off effects. The findings suggest that the usefulness of the ECERS-R for discerning associations with children's outcome may be limited to certain score ranges or subscales.

1. Introduction

During the 1990s, there was an uptick in attention paid to the quality of the care and education that young children experienced in their child care settings in the United States. This attention was driven in part by several multi-state studies that measured preschool quality using the Early Childhood Environment Rating Scale (ECERS; Harms, Clifford, & Cryer, 1980), and reported a national child care quality crisis, particularly for lower-income children (Kagan & Cohen, 1997; Helburn et al., 1995; Loeb, Fuller, Kagan, Carrol, & Carroll, 2004; Whitebook, Phillips, & Howes, 1990). This body of research also demonstrated weak, but positive, associations between the quality of preschool classrooms, as measured by the ECERS, and a number of developmental benefits for preschool-aged children (Peisner-Feinberg & Burchinal, 1997). Preschool quality, as measured by the ECERS, was also shown to be positively associated with children's academic achievement at the early elementary grades (Peisner-Feinberg et al., 2001).

As a result, states began developing child care accountability and quality improvement initiatives, many of which were undergirded by the ECERS-R (the revised version of the ECERS). Presently, quality rating and improvement systems (QRIS) are the most prominent early care and education (ECE) reform effort in the United States, now being implemented in 41 states (Tout et al., 2010), and featured as a required initiative in the federal Race-to-the-Top Early Learning Challenge grants (U.S. Department of Education, 2016). QRIS establish ECE program, classroom, and practitioner quality standards, set thresholds or quality levels on these standards, and measure and monitor the extent to which classrooms meet quality levels. QRIS then provide an overall, summary program quality rating made available to families to assist in their ECE decision-making. Although all states construct their QRIS differently, approximately two thirds of states currently use the ECERS-R as part of their QRIS (QRIS Compendium, n.d.; Administration for Children and Families, 2013).

Currently, 37 states' QRIS attach financial incentives to a program's overall quality rating, the majority of which includes the ECERS-R as a key component (National Center on Early Childhood Quality Assurance, 2017). These financial incentives can include awarding bonuses to teachers based on different thresholds that they have met on the classroom assessment, awarding different levels of payment for children receiving child care subsidies based on a program's rating, and providing programs that meet a certain threshold of quality with improvement grants (Hamilton, Bates, Mitchell, & Workman, 2015; Mitchell, 2012; QRIS Compendium, n.d.). In some states, families are

---

* Corresponding author.
  E-mail address: setodji@rand.org (C.M. Setodji).

also awarded different levels of tuition support based on the rating of the preschool classroom they selected for their child (Schaack, Tarrant, Boller, & Tout, 2012). States provide these incentives tied to higher ECERS-R scores under the assumption that as classrooms meet higher thresholds of quality on the measure, better child outcomes will follow (Zellman & Perlman, 2008).

Yet the ECERS was not originally developed for such high stakes purposes but instead was developed in 1980 as a checklist to help ECE programs prepare for the National Association of the Education of Young Children (NAEYC) Program Accreditation (Frank Porter Graham Child Development Institute, 2003). The definition of quality adopted by the ECERS-R is thus consistent with both the NAEYC program accreditation standards as well as with the Child Development Associate requirements that focuses on the professional knowledge teachers need to facilitate high quality classrooms. Revised in 1998, the ECERS-R is currently constructed with 43 items organized into seven subscales, including Space and Furnishings (8 items), Personal Care Routines (6 items), Facilitation of Children's Language-Reasoning (4 items), Learning Activities (10 items), Teacher Interactions (5 items), Program Structure (4 items), and Supports for Parents and Staff (6 items). Two appealing features of the ECERS-R that contributes to it being one of the most widely used ECE quality measures in state policy are its ease of use and its comprehensiveness. First, the ECERS-R provides specific information about dimensions of quality in which programs are low scoring, and QRIS coaches can use this information to target quality improvement efforts, grants, and professional development to low scoring areas to help programs improve their ECERS-R scores. Second, as a global measure of quality, it is comprehensive in scope, and assesses both structural and process aspects of quality.

Recently, researchers, practitioners, and QRIS designers have begun to raise concerns about the ECERS-R, especially for use in high stakes contexts (Gordon, Fujimoto, Kastner, Korenman, & Abner, 2013; Zellman, Perlman, Le, & Setodji, 2008). Its critics believe that the ECERS-R does not focus enough on aspects of teaching processes that promote conceptual development across learning domains, or on the types of caregiving behaviors that promote secure and trusting relationships with children that facilitate adaptive social-emotional functioning that enable children to engage in learning (Sabol & Pianta, 2014). Instead, those critics believe that the ECERS-R places too much emphasis on environmental quality (Cassidy, Hestenes, Hegde, Hestenes, & Mims, 2005). Concerns have also been raised about the empirical evidence available on the ECERS-R to justify its use in high-stakes contexts (Gordon et al., 2013). Developed as a self-assessment measure intended to provide feedback to programs about their quality, it is unknown whether the ECERS-R can be used to support high-stakes decisions.

1.1. Purpose of this study

The goal of this study is to evaluate the potential utility of the ECERS-R by examining the associations between ECERS-R scores and children's cognitive and social-emotional outcomes, with an emphasis on assessing whether the associations are limited to certain ECERS-R score ranges or on particular ECERS-R subscales. More specifically, we address the following research questions:

1. Are there thresholds on the ECERS-R that are related to preschool children's cognitive and social-emotional outcomes?
2. Do the thresholds change when considering the ECERS-R subscales compared with the total ECERS-R score?

1.2. Psychometric properties of the ECERS-R

In light of the fact that the use of the ECERS-R has expanded and is now being widely used as an accountability measure on which high-stakes decisions are made, there is a growing body of literature that has begun to examine the psychometric properties of the ECERS-R. A number of studies, for example, have subjected the measure to factor analytic techniques to examine its dimensionality. Several studies have observed the ECERS-R to be unidimensional (Holloway, Kagan, Fuller, Tsou, & Carroll, 2001; Perlman, Zellman, & Le, 2004), or two dimensional consisting of factors that tap into the physical environment/ materials and teacher interactions (Cassidy et al., 2005; Sakai, Whitebook, Wishards, & Howes, 2003), but no studies have found evidence of the seven scales described in the ECERS-R.

Using item response theory, Gordon et al. (2013) also demonstrated evidence of individual item multidimensionality, resulting in rating category disorder on 32 of the 36 items they examined. Within the ECERS-R, individual items are composed of multiple binary indicators. Training procedures for the ECERS-R indicate that raters should stop scoring an item once a classroom has not met a particular indicator. As a result of this scoring convention, when an observer assigned a low score on one of the items due to a classroom not meeting a particular indicator on the item, the observer often missed scoring other indicators within the item as higher quality. This may mean that the level of quality needed for classrooms to earn a score of 6, for example, could be less than the level of quality needed to earn a score of 5 (Gordon et al., 2013). This type of research has raised some concerns about how the ECERS-R is constructed.

A small body of research has also examined the association of the ECERS-R against other measures of developmentally appropriate practices, classroom structural quality indices, and children's developmental outcomes. In this research, small to moderate correlations with overall ECERS-R scores have been found with measures of instructional quality such as the Classroom Scoring and Assessment System (CLASS; Mashburn et al., 2008) and the Early Language and Literacy Classroom Observation (ELLCO; Smith & Dickinson, 2002). In a recent meta-analysis of child care quality studies conducted in both the U.S. and in international settings, Vermeer et al. (2016) found strong, positive correlations between the ECERS-R and the teacher sensitivity sub-scale on the Caregiver Interaction Scale (Arnett, 1989). However, with respect to classroom structural quality, weak correlations have been detected between the ECERS-R and teachers' education level and classroom ratios (Early et al., 2006; Gordon et al., 2013; Zellman et al., 2008). No significant associations were detected between ECERS-R scores and classroom group sizes in a meta-analysis using 17 studies (Vermeer et al., 2016). Mixed evidence has also been found when examining the associations between the ECERS-R and developmental outcomes for young children. Some studies have reported weak but positive linear associations with preschoolers' receptive and expressive language skills, applied problem-solving skills, and some indices of social-emotional development (Early et al., 2006; Mashburn et al., 2008). Other studies have failed to find significant associations between the ECERS-R and these dimensions of children's developmental outcomes (Sabol & Pianta, 2014; Zellman et al., 2008).

1.3. Quality thresholds on the ECERS-R

One aspect of the ECERS-R that has been understudied is the existence of thresholds or cut-points along the ECERS-R that may differentiate among the different levels of children's developmental functioning. As discussed by Burchinal, Vandergrift, Pianta, and Mashburn (2010), thresholds are of particular interest to researchers and policymakers because they can inform efficient allocation of resources. They note:

> Most of the literature has examined linear associations, yielding findings that higher quality is better and lower quality is worse (Vandell, 2004), but identification of thresholds in the association between quality and child outcomes has been a goal of researchers and policy makers for several reasons. A primary goal has been to identify levels in the association between quality and child