A comparative analysis of the *Early Childhood Environment Rating Scale–Revised* and *Early Childhood Environment Rating Scale, Third Edition*

Jennifer Neitzel* (iD)

Diane Early
The University of North Carolina at Chapel Hill, USA

John Sideris
University of Southern California, USA

Doré LaForrett
The University of North Carolina at Chapel Hill, USA

Michael B Abel
National Louis University, USA

Margaret Soli
Child Trends, USA

Dawn L Davidson

Amanda Haboush-Deloye
University of Nevada, Las Vegas, USA

Linda L Hestenes
University of North Carolina at Greensboro, USA

Denise Jenson
Georgia Department of Early Care and Learning, USA

Cindy Johnson
Nevada Department of Education, USA

Jennifer Kalas
Nevada Silver State Stars QRIS, USA

Angela Mamrak
The Pennsylvania Key, USA

Marie L Masterson
National Louis University, USA

Sharon U Mims
University of North Carolina at Greensboro, USA

Patti Oya
Nevada Department of Education, USA

Bobbi Philson

Megan Showalter
The Pennsylvania Key, USA

Mallory Warner-Richter
Child Trends, USA

Jill Kortright Wood
The Pennsylvania Key, USA

*After the first four authors, all authors contributed equally to this project and are listed in alphabetical order.

Corresponding author:
Jennifer Neitzel, Frank Porter Graham Child Development Institute (FPG CDI), The University of North Carolina at Chapel Hill, CB #8180, Chapel Hill, NC 27599-8180, USA.
Email: jen.neitzel@unc.edu

Abstract

The Early Childhood Environment Rating Scales, including the *Early Childhood Environment Rating Scale–Revised* (Harms et al., 2005) and the *Early Childhood Environment Rating Scale, Third Edition* (Harms et al., 2015) are the most widely used observational assessments in early childhood learning environments. The most recent version of the scale addresses some of the criticisms in the research literature, particularly related to the organization of the Scale and the standard scoring procedures. In the current study, we explore the relationship between the two scales. Specifically, we evaluated the correlations between the *Early Childhood Environment Rating Scale–Revised* and the *Early Childhood Environment Rating Scale, Third Edition*, as well as the differences between the overall scores, individual items, and subscales. Implications for practice and future research are also provided.

Keywords

Quality in the field of early childhood education has been a key focus of research and program improvement efforts over the past 20 years in the United States and worldwide. Researchers, advocates, and practitioners generally agree that the positive or negative interactions and transactions that young children experience with teachers, materials, and peers are the best way to define and measure quality within a classroom or program (Denny et al., 2012; Early et al., 2007; Howes et al., 2008; Pianta et al., 2005). Questions about the relationship between program quality and children's developmental outcomes have resulted in a proliferation of research related to this topic. Findings from these studies often support the notion that higher quality programs lead to better outcomes for young children both in the short term and long term (Auger et al., 2014; Barnett et al., 2007; Cryer et al., 2003; La Paro et al., 2004; Mashburn et al., 2008; Shonkoff and Phillips, 2000; Vandell, 2004); however, these associations are inconsistent and small in size (Burchinal et al., 2011).

The evidence for a link between quality and children's outcomes has led to an increased emphasis on improving the quality of licensed child care and publicly funded pre-kindergarten programs across the United States. Several state and national initiatives have been developed to boost quality including the development of learning standards/guidelines for young children (Administration for Children and Families, 2014; Daily et al., 2010; QRIS National Learning Network, 2011; U.S. Department of Education, 2014), Quality Rating and Improvement Systems (QRIS; Administration for Children and Families, 2016), Child Care and Development Block Grant Act (2014), and Race to the Top-Early Learning Challenge grants (RTT-ELC; U.S. Department of Education, 2013). A focus of all of these initiatives is to raise quality through program evaluation; increase parents' access to information about quality; and implement supports and requirements for increasing program quality (Denny et al., 2012; Zaslow et al., 2016).

One of the most commonly used systems for measuring quality is the Environment Rating Scales (ERS), which includes tools for measuring quality in center-based classrooms serving infants and toddlers, preschoolers, and school-aged children, as well as family child care. The ERS tool for measuring quality in preschool classrooms, the Early Childhood Environment Rating Scale (ECERS), was recently revised and now many researchers, professional development providers, and programs have switched or are contemplating a switch to the new version. The current article compares this latest version, the *Early Childhood Environment Rating Scale, Third Edition* (ECERS-3; Harms et al., 2015) to its predecessor, the *Early Childhood Environment Rating*

Scale–Revised (ECERS-R; Harms et al., 1998), as a means of aiding users of these tools in making the transition from the older to the newer version.

Early Childhood Environment Rating Scale–Revised

Since its publication in the late 1990s, the ECERS-R has been one of the most widely used program quality measures in the world. It serves as the foundation of quality measurement in nearly every QRIS across the United States (Administration for Children and Families, 2014; Tout et al., 2010) and has been used in many national studies of early childhood, including Head Start Family and Child Experiences Survey (FACES; Moiduddin et al., 2012) and the Early Childhood Longitudinal Study-Birth Cohort (ECLS-B; National Center for Education Statistics, 2016). Researchers also have employed the ECERS-R to evaluate state-funded pre-kindergarten programs (Barnett et al., 2007; Early et al., 2007; Gormley et al., 2005).

The ECERS-R is organized into seven subscales: Space and Furnishings, Personal Care Routines, Language–Reasoning, Activities, Interaction, Program Structure, and Parents and Staff. However, few users score the Parent and Staff items and most users calculate a Total Score, which is the arithmetic mean of all items scored, rather than subscale scores. The sub-scales and items are organized to promote ease of observation by allowing similar information to be collected across settings and raters (Cryer et al., 2003). The ECERS-R is generally admin-istered over a 3-hour period in which observers respond to hundreds of yes/no indicators. Following the observation, staff are asked a variety of questions about activities that typically occur but not have been observed, and their responses are used to answer any of the yes/no indi-cators that remain unscored. The pattern of responses to the indicators determines the scores on 37[1] 7-point items. According to the ECERS-R authors, 1 indicates *inadequate* quality, 3 indi-cates *minimal* quality, 5 indicates *good* quality, and 7 indicates *excellent* quality. When applying the standard scoring procedures, observers stop scoring yes/no indicators for each item when a score has been obtained.

Previous research suggests that the ECERS-R demonstrates adequate reliability and validity to be used in both quality assurance and research, including inter-rater reliability, and convergent validity (Pianta et al., 2008); however, the associations between program quality and child out-comes are modest at best (Burchinal et al., 2002; Howes et al., 2008; Love et al., 2005). Concern over the limited predictive validity, coupled with the widespread use of the ECERS-R, has led to growing concern from researchers and policymakers in recent years (Gordon et al., 2013).

Early Childhood Environment Rating Scale, Third Edition

According to the authors of the ECERS-3 (Harms et al., 2015), the revisions were intended to address shortcomings identified within the research literature and to incorporate the latest research and thinking about quality in early childhood settings, which includes an increased focus on teacher behaviors particularly related to language/literacy and math (Gordon et al., 2013). The newly revised ECERS-3 shares many common features with the ECERS-R. For example, both tools cover the broad range of children's developmental needs, including cognitive, social-emotional, physical, health, and safety. The general structure remains the same, with yes/no indicators answered to derive scores on 7-point items. Examples of items that appear in both the ECERS-R and the ECERS-3 include Health Practices, Space for Gross Motor Play, and Staff–Child Interactions.

Despite these similarities, the changes to the tool are substantial, with the ECERS-3 placing much more emphasis on the role of the teacher in helping children develop both cognitive and

social skills, with somewhat less emphasis on provision of materials. The differences between the two tools fall into three broad categories: (1) additional/refined items; (2) additional/refined indicators; and (3) changes to the procedures for conducting the observation and scoring. New items were added to strengthen the emphasis on the role of the teachers, the importance of high-quality teacher–child interactions, and individualized teaching and learning, as supported by research (Chien et al., 2010; Pianta et al., 2005) and widely accepted best practice (Epstein, 1993). For instance, there are five new 7-point items regarding language and literacy (e.g. Helping Children Expand Vocabulary, Encouraging Children to Use Language), and each of them focuses on the role of teachers in promoting these important early skills. Likewise, whereas the ECERS-R had a single 7-point item regarding Math/Number, the revised ECERS-3 has three 7-point math items: Math Materials and Activities, Math in Daily Events, and Understanding Written Numbers.

Yes/no indicators have been added to some items to address evidence that additional information was needed to distinguish among the highest levels of quality (Gordon et al., 2013) and to further strengthen the emphasis on the role of the teacher. For example, in the Fine Motor item, two new indicators have been added at the excellent level regarding how staff members work with children to extend and expand children's experiences with fine motor materials, making the distinction between the 5 (good) and the 7 (excellent) more pronounced. In addition, for items that consistently received very low scores on the ECERS-R, such as those regarding Personal Care Routines like Toileting and Health Practices, the ECERS-3 incorporates additional indicators and revised definitions at the lower levels of quality to better distinguish between truly inadequate practice and occasional lapses of supervision or attention.

Finally, some definitions, rules, and procedures have been revised. For instance, the ECERS-R often required that materials and activities be available for a "substantial portion of the day." That phrase was defined as one-third of the time the children were in attendance, but it was difficult for raters to gauge how time was used outside of the observation period. Further, a review of states' early learning guidelines and standards revealed that regulations about instruction, rest, outside time, and meals made it impossible for some programs to make some materials and activities available for one-third of the time, decreasing the accuracy of their scores (Harms et al., 2015). The ECERS-3 has addressed these problems by removing "substantial portion of the day" and providing more simplified and targeted time requirements. Another change to procedures is that there is no longer a teacher interview component. All items are scored based solely on what is observed during the 3 hours. This was made possible, in part, by the elimination of the Parents and Staff subscale, which was scored mainly based on teacher and staff reports. All observations now last exactly 3 hours (rather than 3 or more hours) and the time requirements specified within the indicators are tied to that observation length.

Another significant revision of the ECERS-R was related to the standard scoring procedures. With the guidelines outlined in the ECERS-R, there was no need to score the higher level indicators within an item if the lower level indicators were not met. Observers only answered as many yes/no indicators as needed to derive the scores on the 7-point items. This "stop scoring" approach led to loss of information about some aspects of program quality, particularly at the upper end of the Scale. For example, if an observer awarded a score of "2" on the Nature/Science item, no indicators above a level 3 would be scored, even if a program was implementing some of the upper level practices. The ECERS-3 incorporates new scoring procedures in which all indicators are scored either yes or no, regardless of whether a threshold needed to score the item has been reached. Scoring all indicators provides more information to support program development and improvement. Under the ECERS-R guidelines, data were often missing from the upper end of the scoring continuum. The revised scoring

procedures offer programs and technical assistance providers with more nuanced information about strengths and areas for improvement within individual classrooms.

In the current study, we focused on the relationship between the ECERS-R and ECERS-3. Specifically, we used secondary data to conduct a comparative analysis of the two scales and asked the following research questions:

1. Is there a systematic difference between ECERS-R and ECERS-3 scores, such that one is typically higher than the other by a fairly stable amount?
2. What is the correlation between the ECERS-R and the ECERS-3? (A high correlation would indicate they are measuring largely the same underlying construct.)
3. How similar/different are individual items and subscales on the Scales (underlying constructs they are measuring) and how correlated are similar items from the two tools?

This study provides insights into the difference between the two scales, which can help guide users as they make decisions about identifying appropriate measures for research, evaluation, and rating systems such as QRIS.

Method

Sample

The current secondary data analysis includes 225 observations in which ECERS-R and ECERS-3 were conducted by different individuals in the same classroom on the same day. These data were collected in six states that were transitioning or were considering transitioning from using the ECERS-R to the ECERS-3 for their state QRIS. The six states were Colorado, Georgia, Nevada, North Carolina, Pennsylvania, and Vermont. Additional details about each state's sample and data collection can be found in Table 1.

Colorado (n = 37). ECERS-R and ECERS-3 data were collected to determine whether a switch from ECERS-R to ECERS-3 was warranted for Colorado's QRIS, Colorado Shines. Programs that received an automatic rating in Colorado Shines were eligible to participate. They included 43 percent Head Start programs and 57 percent programs that were accredited by an approved national accrediting body (i.e. National Association for the Education of Young Children (NAEYC), American Montessori Society, Association of Christian Schools International, National Early Childhood Program). Participating programs that received an automatic rating were selected at random, with some stratification by geography. The goal was to recruit 75 percent of the sample from programs located within an hour radius of the Denver metro area and the remaining 25 percent from within the state. Recruitment information was sent to 82 Head Start programs and 100 accredited programs, and 16 and 21 agreed to participate, respectively, resulting in an overall response rate of 20 percent. One preschool classroom in each program was selected at random for participation.

Georgia (n = 49). In Georgia, the QRIS, called Quality Rated, was in the process of transitioning from ECERS-R to ECERS-3, and these data were collected to inform that transition. Most of the programs included in the sample were being evaluated to obtain their star rating. Within programs, one preschool classroom was selected at random for participation. Selected classrooms included both licensed child care and state pre-K. The sample include 67 percent licensed child care centers and 33 percent state pre-K classrooms.

Nevada (n = 15). Data were collected as part of Nevada's decision-making process to determine whether they should switch from using the ECERS-R to the ECERS-3 within their QRIS, Nevada Silver State Stars. Classrooms included 13 percent Head Start, 73 percent school district preschool programs, and 13 percent charter schools that were already scheduled for an assessment (either ECERS-3 or ECERS-R). The sites did not have a choice about participation in the study because all QRIS sites are required to allow assessors to practice or collect data as needed.

North Carolina (n = 23). Data were collected as part of a larger descriptive study comparing the ECERS-3 and ECERS-R to examine the similarities and differences between the two versions of the Scale to inform decisions about possible adoption of the ECERS-3 by other states.

Programs were selected based on their upcoming participation in a scheduled assessment for licensing. During the time period of the study (April–September 2015), 694 child care centers were scheduled for ERS assessments or had submitted pending assessment requests and all were invited to take part in the comparison study. Of those, 105 did take part in the study, but only 23 classrooms agreed to have the ECERS-R and ECERS-3 completed at the same time, which made them eligible for participation in the current analyses.

Pennsylvania (n = 51). In Pennsylvania, data were collected to compare the results of the ECERS-R with the results of the ECERS-3 while planning how to incorporate the use of the new tool within their QRIS. Most programs included in the sample were due to be assessed with the ECERS-R to renew their STAR level in the state QRIS during the months the study was conducted and participation was not optional for those programs. Ten sites in the sample did not need an assessment for their QRIS status, but volunteered to have the comparability visit. Within programs, one-third of all eligible preschool classrooms were selected at random for participation. For all but one program, this resulted in a single classroom being selected; in one program, two classrooms participated. The final sample included 4 percent Head Start classrooms within a licensed child care center, 20 percent state-funded pre-K classrooms within a licensed child care center, and 76 percent regular licensed child care (i.e. not Head Start nor state-funded pre-K).

Vermont (n = 50). These data were collected as part of Vermont's QRIS (STARS) Validation study. Programs that received a level 4 or 5 rating were eligible to participate in the ECERS-R/ECERS-3 sub-study. Participating programs were selected at random, with some stratification by geography. Programs were sorted into groups across four tiers of counties. These groups were then contacted in waves for equal distribution. Recruitment information was sent to 169 programs and 50 (30%) agreed to participate. One ECERS-R/ECERS-3 observation was completed in one classroom per program, selected at random. The final sample included 71 percent child care centers, 2 percent Head Start classrooms, and 26 percent school-based pre-K.

Inter-rater reliability

The states had slightly different procedures for training data collectors and testing their inter-rater reliability, but all states followed guidelines similar to those endorsed by the Environment Rating Scales Institute (2018). In all states, after each inter-rater reliability visit, the observers met to determine consensus scores (the group's final determination of the correct score for each item). The percent of items on which observers' scores were within 1-scale point of the consensus score was considered their reliability score for that visit. All states required data collectors to attain 85 percent reliability during between three and five visits in order to be considered reliable and able to collect data independently.

Table 1. Sample description and reliability procedures.

State	n	Program types	Response rate	ECERS-R inter-rater reliability standards	ECERS-3 inter-rater reliability standards[a]
CO	37	43% Head start 57% accredited	20%	85% across three observations	85% across three observations
GA	49	67% licensed child care 33% state pre-K	N/A[b]	Average of 85% across three visits	Average of 85% across three visits
NV	15	13% Head Start 73% school district preschool program 13% charter school	N/A[c]	Anchor: 90% with ERSI, checked annually Lead Assessors: 90% with Anchor, checked monthly Data collectors 85% with state anchor or primary assessor over three most recent observations, checked monthly	Anchor: 90% with ERSI, checked annually Lead Assessors: 90% with Anchor, checked monthly Data collectors 85% with state anchor or primary assessor over three most recent observations, checked monthly
NC	23	52% for profit centers 13% non-profit centers 35% public programs (school, university, or Head Start)	3%	85% across four observations	Anchors: 90% with ERSI Data collectors: 85% across four observations
PA	51	76% licensed child care only 4% Head Start within a licensed child care center 20% state-funded pre-K within a licensed child care center	60%	New assessor: 85% on five visits with a seasoned assessor	New assessor: 85% on five visits with a seasoned assessor + two visits at 85% with state gold standard or ECERS-3 Anchor Seasoned assessor (already ECERS-R reliable): 85% for two visits at 85% with state gold standard or ECERS-3 Anchor
VT	50	71% child care centers 2% Head Start 26% school-based	30%	Average of 85% across three observations	Data collectors: 85% across three observations

ECERS-R: *Early Childhood Environment Rating Scale–Revised*; ECERS-3: *Early Childhood Environment Rating Scale, Third Edition*; CO: Colorado; GA: Georgia; NV: Nevada; NC: North Carolina; PA: Pennsylvania; VT: Vermont; ERSI: Environment Rating Scales Institute.
[a]In all states, the values noted here refer refer the percentage of the individual's item scores that were within 1-scale point of the consensus score for that item.
[b]In Georgia, these observations were conducted as part of the regular observation for participation in Quality Rated. Participation in Quality Rated is voluntary, but allowing the ECERS-R/ECERS-3 observations was not optional for those who had applied for Quality Rated and were selected.
[c]In Nevada, the sites did not have a choice about participation in the study because all QRIS sites are required to allow assessors to practice or collect data as needed.

Results

Systematic difference between the ECERS-R and ECERS-3

Basic descriptive statistics for Total Scores of both versions of the Scale, as well as the subscale scores, are provided in Table 2. A key question of this study was to determine whether ECERS-R and ECERS-3 scores differed systematically such that scores for one were consistently higher or lower than the other. We conducted a series of paired t-tests comparing the mean total and subscale scores of the ECERS-R and ECERS-3. The findings of these tests indicate that the Total Score and all six subscale scores are significantly different from their counterparts. Scores were significantly higher for ECERS-R Total Score and each of the subscale scores, with the exception of Personal Care Routines, where the ECERS-3 score was significantly higher. The largest differences were between the Total Scores ($t(224)=-15.04$, $p<.0001$), the Language subscale ($t(224)=-17.56$, $p<.0001$), and the Activities subscale ($t(224)=-22.62$, $p<.0001$).

In addition, we wanted to understand the magnitude of the differences in Total Scores. When the ECERS-R Total score was subtracted from the ECERS-3 Total Score, the average difference was less than one point ($M=.83$, $SD=.83$). Most classrooms obtained a lower Total Score on the ECERS-3 than on the ECERS-R (85%).

Correlation between the ECERS-R and ECERS-3

Table 2 also provides the correlations between the ECERS-R and ECERS-3 Total Scores, as well as the correlations between the individual subscales. The correlation between the ECERS-R and ECERS-3 was positive and modest ($r=.51$, $p<.001$), indicating that in general programs that score higher on one version also score higher on the other, but that there is not a strong linear association between ECERS-R and ECERS-3 that would allow for easy translation between instruments. There was wide variation among the correlations at the subscale level. Although the Scale authors eliminated two items from the Personal Care subscale for the ECERS-3, scores on that subscale demonstrated the strongest correlation between the two versions ($r=.62$, $p<.001$). The Scale authors also made significant changes to both the number and wording of items for the Program Structure subscale, likely contributing to a weaker correlation ($r=.30$, $p<.001$).

One concern we had with respect to these findings is that the magnitude of the association between the ECERS-R and ECERS-3 may vary by state because there were significant differences in total and subscale ECERS scores among states in the sample. To confirm that the correlation did not vary systematically by state, we used analysis of variance and included the interactions between state and ECERS-R score in predicting ECERS-3 scores. The results indicated that state was not significantly associated with the magnitude of association between the ECERS-R and ECERS-3 (see Table 3).

Similarities and differences between individual items

A final undertaking of the current study was to determine similarities and differences between individual items on the two Scales. Table 4 provides means, standard deviations, *t*-tests, and correlations between items on the ECERS-R and ECERS-3. We elected to compare items whenever they had similar names, despite knowing that the indicators within the items had been substantially changed in some cases. We did this because the similar item names indicated that the authors had intended for them to measure similar constructs and because users of the Scales are likely to assume that their content overlaps. Also, when a single item on the ECERS-R had become multiple items on the

Table 2. Descriptive statistics, paired *t*-tests, and correlations for ECERS-R/ECERS-3 total score and subscale scores (*n* = 225).

Subscale	ECERS-R		ECERS-3		Mean difference	t	r
	Mean (*SD*)	Range	Mean (*SD*)	Range			
Total score	4.42 (.86)	2.08–6.43	3.59 (.82)	1.47–5.77	.83	−15.04	.51
Space and furnishings	4.14 (.98)	2.00–6.63	3.76 (.77)	1.43–6.57	.38	−7.18	.42
Personal care routines	2.99 (1.05)	1.17–6.17	3.15 (.96)	1.25–6.25	−.16	2.78	.62
Language[a]	4.93 (.97)	2.50–7.00	3.73 (1.16)	1.00–6.60	1.20	−17.56	.50
Activities[b]	4.57 (1.04)	2.11–7.00	3.02 (.89)	1.27–6.18	1.55	−22.62	.45
Interaction	5.16 (1.35)	1.50–7.00	4.52 (1.32)	1.00–6.80	.64	−9.39	.59
Program structure	4.61 (1.30)	1.67–7.00	4.01 (1.40)	1.00–7.00	.60	−5.91	.30

ECERS-R: *Early Childhood Environment Rating Scale–Revised*; ECERS-3: *Early Childhood Environment Rating Scale, Third Edition*; SD: standard deviation.
p values for all *t*s and *r*s are < .001, except Personal Care Routines (*p* < .01)
[a]In ECERS-R, this subscale is called *Language–Reasoning*; in ECERS-3, it is called *Language and Literacy*.
[b]In ECERS-R, this subscale is called *Activities*; in ECERS-3 it is called *Leaning Activities*.

ECERS-3, such as the change from Math/Number to Math Materials and Activities, Math in Daily Events, and Understanding Written Numbers, we compared the single ECERS-R item with each of the individual ECERS-3 items.

The *t*-tests indicated that there were significant differences between all of the individual items on the Scales, except for Furnishings (item 2 on both scales) and Health Practices (item 13 on the ECERS-R and item 10 on the ECERS-3). The items with the largest mean differences included Communication/Vocabulary (item 16 on ECERS-R and 12 on ECERS-3; $t(224) = 19.99, p < .0001$), Music/Movement (item 21 on ECERS-R and 19 on ECERS-3 ($t(224) = 15.46, p < .0001$), Blocks (item 22 on ECERS-R and 20 on ECERS-3; $t(224) = 15.68, p < .0001$), and Nature/Science (item 25 on ECERS-R and 22 on ECERS-3; $t(224) = 15.01, p < .0001$). In addition, there were large differences between the Math item on the ECERS-R (item 26) and Math Materials and Activities (item 23; $t(224) = 16.21, p < .0001$) and Understanding Written Numbers (item 25; $t(224) = 21.15, p < .0001$) on the ECERS-3.

Discussion

The overall findings from this study comparing two versions of the *ECERS* seem to indicate that the ECERS-R and ECERS-3 are two distinct quality measurement tools, rather than one instrument measuring the same features within learning environments. The differences between the two scales, particularly related to the emphasis on teacher–child interactions within the ECERS-3, reflect an evolution in the concept of quality within the field of early childhood education. When the ECERS was first published in the 1980s, quality was generally defined in terms of structural indicators (e.g. staff:child ratio, provision of materials; Dahlberg et al., 2006). Since this time, quality has been re-conceptualized to emphasize the importance of process quality that views child development within the context of the instructional and personal interactions that take place throughout the course of a school day. This post-modern view of program quality suggests that the notion of quality is socially constructed and should be considered from the standpoint of current societal norms and values

Table 3. Analysis of variance predicting ECERS-3 Total Score (*n* = 225).

	df	*F*	ω^2	*p*
ECERS-R Total Score	1	107.28	.23	<.001
State	5	2.37	.01	.040
ECERS-R × State Interaction	5	.95	.00	.45

ECERS-R: *Early Childhood Environment Rating Scale–Revised.*

regarding how children grow and develop (Dahlberg et al., 2006). As such, this outlook provides a better understanding about the continued revision of the ECERS scale as well as why the ECERS-R and ECERS-3 are best understood as two distinct measures—a key finding of the current study.

Our results suggest that the ECERS-R and ECERS-3 were only modestly correlated which does not allow for easy translation between the instruments. For example, the correlation between the two versions of the ECERS is about the same as between either of them, and the *Classroom Assessment Scoring System* (CLASS; Early et al., 2018; Pianta et al., 2008), another widely used early childhood classroom quality measure. Thus, while the ECERS-R and ECERS-3 are measuring some common characteristics, they are distinct tools that cannot be used interchangeably. Assessors cannot simply add or subtract a constant amount from the Total Score on one version of the ECERS and assume that it is a good estimate of the score that program would have received on the other. Instead, users should carefully consider the aspects of quality they are seeking to measure and determine which tool best matches their goals.

One potential reason for the modest relationship between the two measures is that significant changes were made to both the number and wording of items in the ECERS-3 Program Structure subscale. Items within this subscale provided more specificity about the content, level of child engagement, and teacher interactions within whole group (e.g. "All children in the group are actively engaged," "Staff use group time to introduce children to meaningful ideas in which children are interested") and free play activities (e.g. "Staff use a wide variety of words to expand children's knowledge during free play activities," "Staff interact positively with children during free play"). A greater emphasis on teacher behaviors reflects the changing discourse about program quality and what we now consider to be important standards or performance indicators within the field of early childhood. Within the past two decades, there has been a shift from ensuring that children have access to developmentally appropriate materials to how teachers facilitate the learning for young children throughout the day (Dahlberg et al., 2006). The fact that the correlation between the two versions of the Program Structure subscale is only .30 (the lowest of any subscale) supports the idea that changes to this subscale are an important factor driving the overall difference. An additional goal of this study was to develop a greater understanding about the systematic differences between the ECERS-R and ECERS-3. Our findings indicate that scores were significantly higher for ECERS-R Total Score and each of the subscale scores, with the exception of Personal Care Routines, perhaps related to the similarity of the remaining items within the two measures. It is not surprising, given the revisions to the Scale, that ECERS-R Total Score and subscale scores were generally higher, particularly the Language and Activities' subscales. We speculate that these discrepant scores also were mostly due to the significant changes to these items within the ECERS-3, including additional and more difficult indicators that were intended to be more related to teacher behavior rather than provision of materials. Examples of new indicators from the Language subscale include: "Staff generally use a wide range of words to specify more exactly what they are talking about," "Staff–child conversations go beyond classroom activities and materials," and "Staff add information and ideas in order to expand children's understanding

Table 4. Paired *t*-tests and correlations for ECERS-R and ECERS-3 items (*n* = 225)[a].

ECERS-R item	ECERS-3 item	ECERS-R mean (SD)	ECERS-3 mean (SD)	t	r
1. Indoor space	1. Indoor space	4.10 (1.45)	4.29 (1.55)	−1.68	.36
2. Furniture for routine care, play, learning	2. Furnishings for care, play, learning	4.00 (2.02)	3.96 (1.12)	.31	.36
4. Room arrangement for play	3. Room arrangement for play, learning	4.67 (1.72)	3.98 (1.49)	4.92***	.13
5. Space for privacy	4. Space for privacy	4.90 (1.87)	4.49 (1.60)	2.86**	.24
6. Child-related display	5. Child-related display	4.67 (1.51)	3.14 (1.42)	13.34***	.31
7. Space for gross motor play	6. Space for gross motor play	2.33 (1.45)	3.30 (1.34)	−9.07***	.34
8. Gross motor equipment	7. Gross motor equipment	4.47 (2.31)	3.23 (1.87)	8.17***	.41
10. Meals/snacks	8. Meals/snacks	2.36 (1.52)	3.08 (1.21)	−6.53***	.27
12. Toileting/diapering	9. Toileting/diapering	2.19 (1.76)	3.25 (1.35)	−10.05***	.50
13. Health practices	10. Health practices	2.92 (1.84)	2.95 (1.44)	−.21	.32
14. Safety practices	11. Safety practices	1.87 (1.31)	3.32 (1.63)	−14.24***	.47
16. Encouraging children to communicate	12. Helping children expand vocabulary	6.26 (1.21)	3.88 (1.62)	19.99***	.23
15. Books and pictures	14. Staff use of books with children	4.71 (1.40)	3.57 (1.68)	9.15***	.28
15. Books and pictures	15. Encouraging children's use of books	4.71 (1.40)	3.80 (1.64)	7.89***	.37
18. Informal use of language	13. Encouraging children to use language	5.36 (1.53)	4.30 (1.68)	9.81***	.49
19. Fine motor	17. Fine motor	5.08 (1.65)	4.36 (1.69)	6.03***	.42
20. Art	18. Art	4.35 (1.55)	3.48 (1.40)	7.50***	.30
21. Music/movement	19. Music and movement	4.87 (1.84)	3.10 (1.15)	15.46***	.42
22. Blocks	20. Blocks	3.84 (1.20)	2.22 (1.25)	15.68***	.21
24. Dramatic play	21. Dramatic play	4.30 (1.38)	3.10 (1.75)	9.65***	.30
25. Nature/science	22. Nature/science	4.56 (1.71)	2.79 (1.30)	15.01***	.34
26. Math/number	23. Math materials and activities	4.19 (1.69)	2.16 (1.30)	16.21***	.23
26. Math/number	24. Math in daily events	4.19 (1.69)	2.90 (1.49)	9.48***	.18
26. Math/number	25. Understanding written numbers	4.19 (1.69)	1.72 (1.06)	21.15***	.25
28. Promoting acceptance of diversity	26. Promoting acceptance of diversity	5.03 (1.56)	4.18 (1.27)	8.38***	.44
29. Supervision of gross motor activities	28. Supervision of gross motor	4.61 (1.9)	3.98 (1.69)	4.70***	.40
31. Discipline	32. Discipline	5.09 (1.66)	4.28 (1.66)	6.86***	.54
32. Staff–child interactions	30. Staff–child interaction	5.82 (1.89)	5.12 (1.85)	5.55***	.49
33. Interactions among children	31. Peer interaction	5.74 (1.67)	4.50 (1.72)	9.97***	.40
34. Schedule	33. Transitions and waiting times	3.51 (1.71)	3.89 (2.22)	−2.29*	.20
35. Free play	34. Free play	4.74 (2.07)	4.24 (1.50)	3.20**	.16
36. Group time	35. Whole-group activities for play, learning	5.24 (1.71)	3.82 (1.60)	9.76***	.21

ECERS-R: *Early Childhood Environment Rating Scale–Revised*; ECERS-3: *Early Childhood Environment Rating Scale, Third Edition*; *SD*: standard deviation.
[a]*n* for all items was 225 except Group (*n* = 218). The range for all ECERS-R items was 1–7, except items 4, 6, 15, 16, and 20 where it was 2–6. The range for all ECERS-3 variables was 1–7, except item 22 where it was 1–6 and item 28 where it was 2–7.
*p < .05; **p < .01; ***p < .001.

of the meaning of words children use." It may be that these additional items and indicators, particularly at the upper end of the Scale, make it more difficult for classrooms to achieve higher ratings; however, additional research is needed to fully understand the increasingly difficult nature of individual items on the ECERS-3. Again, this emphasis on teacher-related behaviors reflects the evolution of how quality is currently defined within the field of early education.

Finally, we wanted to explore the similarities and differences between individual items on the ECERS-R and ECERS-3. The results from the analyses indicated that there were significant differences between all of the individual items on the Scales, except two. There were not significant differences on the scores for the Furnishings and Health Practices items. These two items underwent minimal changes between versions, so the similarity in their scores are not surprising. Although other notions of quality have evolved over the years, health, safety, and furnishings have remained primary indicators of high-quality programs.

However, for the other items, the Scale authors engaged in a significant revision process which resulted in a greater emphasis on teacher interactions than on provision of materials. All of the ECERS-3 items that were found to differ significantly from the ECERS-R counterparts (i.e. Vocabulary, Music, Blocks, Science) include greater involvement by staff related to interactions with children and facilitation of learning (e.g. "Staff frequently use the opportunities provided by materials, display, activities, or other meaningful experiences to introduce words"; "Staff point out rhyming words in songs, identify sound repetition or do finger plays with children use gestures or actions to act out the meaning of words"; "Staff point out the math concepts that are demonstrated in unit blocks in a way that interests children"; and "Staff initiate activities for measuring, comparing, or sorting nature/science materials"). These changes reflect a growing understanding about what constitutes quality in early childhood education.

Implications for practice and future research

Combined, the findings from the current study indicate that the ECERS-3 should be viewed as a separate instrument that is truly distinct from the ECERS-R, rather than simply a minor update. Individuals seeking to measure early childhood classroom quality should carefully review the two instruments to select the one that best matches their purposes, rather than thinking of them as a single tool. The ECERS-3 has been well received by researchers and practitioners, including the authors of this article, because it reflects the field's increasing understanding of the role of the teacher and holds programs to very high standards. That said, the current analyses indicate that it is quite distinct from its predecessor, so users should think carefully about their goals and values when selecting the appropriate instrument.

The findings from the study also have important implications for future research. For example, future research efforts should focus on replicating the data analyses with representative samples and with samples that intentionally sample from the full range of quality. Findings from such studies would provide more in-depth information about the differences between the two scales. Previous research on the ECERS-R indicates that the Scale contains two factors: *Activities/Materials* and *Language/Interactions* (Cassidy et al., 2005) or *Provisions for Learning* and *Teaching and Interactions* (Clifford et al., 2005). More recently, researchers have found that a four-factor solution offered the best fit for the ECERS-3: *Learning Opportunities, Gross Motor, Teacher Interactions*, and *Math Activities* (Early et al., 2018). The sample size in the current study did not support this level of analysis, but it would be interesting to see if these factors were replicated when the data were collected in the same rooms at the same time.

Additional research also should include child outcome measures to determine differences in the predictive validity between the two scales, which has become increasingly important as the

role of early childhood programs has shifted from primarily providing care to preparing children for school entry (Dahlberg et al., 2006). Past research on the ECERS-R suggests that there have been only small to modest associations between program quality and child outcome measures (Burchinal et al., 2011; Shonkoff and Phillips, 2000; Vandell, 2004; Yoshikawa et al., 2013). More recent research on the ECERS-3 has found that associations with children's outcomes provided some evidence of the tool's predictive validity, but the associations were small and not domain-specific (Early et al., 2018). To fully understand the differences in the predictive validity between the scales, future research should focus on conducting studies in which ECERS-R and ECERS-3 are administered simultaneously in classrooms where child outcomes data are collected.

Another future avenue of research is to place a greater emphasis on understanding development within the context of children's cultures. According to Dahlberg et al. (2006), childhood is a social construction that is best understood in relation to time, place, and children's culture. In addition, they also argue that childhood varies according to children's class, gender, and other socioeconomic conditions. As our understanding of these variables evolves, definitions of quality also will change accordingly. Researchers and theorists should continue to place an emphasis on identifying how child development is influenced by these factors so that we can create program quality indicators that fully reflect the learning styles and needs of all young children.

Limitations

The findings from this study offer important implications for the field; however, they should be viewed within the context of several limitations. First, the sample was one of convenience. The three lead authors approached the participating states knowing that they had collected both ECERS-R and ECERS-3 data to inform practice and policy within their states. Each state was, in essence, conducting its own study with differing procedures for recruitment. Therefore, the overall sample is not representative and there is significant state-to-state variation in the types of programs included, as well as the quality of those programs. After the data were received from the states, they were combined into one data set to complete the analyses for the study. These issues related to the sample, combined with differing procedures for recruitment and classroom selection, limit the generalizability of our findings.

Conclusion

The purpose of this study was to examine the relationship between the ECERS-R and the newly revised ECERS-3. To accomplish this task, we conducted a comparative analysis using secondary data to determine the two versions' relationship to one another as well as the differences between the two scales. The findings from the study offer state QRIS and policymakers important information about the similarities and differences between the two scales and offer some support in making decisions about which Scale is most appropriate for measuring quality within early childhood programs.

Authors' Note

Funding

ORCID iD

Jennifer Neitzel https://orcid.org/0000-0003-0323-0074

Note

1. There are 37 items on the portion of the ECERS-R that is typically scored. If the Parent and Staff items are scored, there are 43 items.

References

Administration for Children Families (2014) National program standards crosswalk tool: quality rating and improvement system resource guide. Available at: https://ecquality.acf.hhs.gov/crosswalk

Administration for Children Families (2016) Quality rating and improvement system resource guide. Available at: https://ecquality.acf.hhs.gov/

Auger A, Farkas G, Burchinal MR, et al. (2014) Preschool center care quality effects on academic achievement: an instrumental variables analysis. *Developmental Psychology* 50(12): 2559–2571.

Barnett WS, Friedmann AH, Boyd JS, et al. (2007) The state of preschool 2007: state preschool yearbook. Available at: http://nieer.org/wp-content/uploads/2016/10/2007yearbook.pdf

Barnett WS, Friedmann-Krauss AH, Gomez R, et al. (2016) The state of preschool 2015: state preschool yearbook. Available at: http://nieer.org/research/state-preschool-2015

Burchinal M, Kainz K and Cai Y (2011) How well do our measures of quality predict child outcomes? A meta-analysis of data from large-scale studies of early childhood settings. In: Zaslow M, Martinez-Beck I, Tout K, et al. (eds) *Quality Measurement in Early Childhood Settings*. Baltimore, MD: Brookes, pp. 11–32.

Burchinal MR, Peisner-Feinberg E, Pianta R, et al. (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.

Cassidy DJ, Hestenes L, Hedge A, et al. (2005) Measurement of quality in preschool child care classrooms: An exploratory and confirmatory analysis of the early childhood environment rating scale-revised. *Early Childhood Research Quarterly* 20: 345–360.

Chien NC, Howes C, Burchinal M, et al. (2010) Children's classroom engagement and school readiness gains in prekindergarten. *Child Development* 81(5): 1534–1549.

Clifford RM, Barbarin O, Chang F, et al. (2005) What is pre-kindergarten? Characteristics of public pre-kindergarten programs. *Applied Developmental Science* 93(3): 126–143.

Cryer D, Harms T and Riley C (2003) *All about the ECERS-R*. Lewisville, NC: Kaplan.

Dahlberg G, Moss P and Pence A (2006) *Beyond Quality in Early Childhood Education and Care: Postmodern Perspectives*. New York: RoutledgeFalmer.

Daily S, Burkhauser M and Halle T (2010) A review of school readiness practices in the states: early learning guidelines and assessments. *Child Trends* 1(3). Available at: https://nces.ed.gov/programs/slds/pdf/10_III_IVb.pdf

Denny JH, Hallam R and Homer K (2012) A multi-instrument examination of preschool classroom quality and the relationship between program, classroom, and teacher characteristics. *Early Education and Development* 23: 678–696.

Early DM, Maxwell KL, Burchinal M, et al. (2007) Teachers' education, classroom quality, and young children's academic skills: results from seven studies of preschool programs. *Child Development* 78(2): 558–580.

Early DM, Sideris J, Neitzel J, et al. (2018) Factor structure and validity of the Early Childhood Environment Rating Scale–Third Edition (ECERS-3). *Early Childhood Research Quarterly* 44(3): 242–256.

Environment Rating Scales Institute (2018) Frequently asked questions. Available at: https://www.ersi.info/faq_training.html#reliability

Epstein A (1993) *Training for Quality: Improving Early Childhood Programs through Systematic Inservice Training*. Ypsilanti, MI: High/Scope Press.

Gordon RA, Fujimoto K, Kaestner R, et al. (2013) An assessment of the validity of the ECERS-R with implications for assessments of child care quality and its relation to child development. *Developmental Psychology* 49(1): 146–160.

Gormley WT, Gayer T, Phillips D, et al. (2005) The effects of universal pre-K on cognitive development. *Developmental Psychology* 41(6): 872–884.

Harms T, Cliffor R and Cryer D (1998) *Early Childhood Environment Rating Scale-Revised*. New York: Teachers College Press.

Harms T, Clifford R and Cryer D (2005) *Early Childhood Environment Rating Scale–Revised*. New York: Teachers College Press.

Harms T, Clifford R and Cryer D (2015) *Early Childhood Environment Rating Scale* (3rd ed.). New York: Teachers College Press.

Howes C, Burchinal M, Pianta R, et al. (2008) Ready to learn? Children's pre-academic achievement in pre-kindergarten programs. *Early Childhood Research Quarterly* 23(1): 27–50.

La Paro KM, Pianta RC and Stuhlman M (2004) The classroom assessment scoring system (CLASS): findings from the prekindergarten year. *Elementary School Journal* 104: 409–426.

Love JM, Kisker EE, Ross C, et al. (2005) The effectiveness of early head start for 3-year-old children and their parents: lessons for policy and programs. *Developmental Psychology* 41: 885–901.

Mashburn AJ, Pianta RC, Hamre BK, et al. (2008) Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development* 79(3): 732–749.

Moiduddin E, Aikens N, Tarullo L, et al. (2012) *Child outcomes and classroom quality in FACES 2009*. OPRE Report Number 2012–37a, September 2012. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families. Available at: http://www.acf.hhs.gov/sites/default/files/opre/faces_2009.pdf

National Center for Education Statistics (2016) Early childhood longitudinal program (ECLS): data collection procedures. Available at: http://nces.ed.gov/ecls/birthdataprocedure.asp

Pianta R, Howes C, Burchinal M, et al. (2005) Features of prekindergarten programs, classrooms, and teachers: do they predict observed classroom quality and child-teacher interactions? *Applied Developmental Science* 9(3): 144–159.

Pianta R, La Paro KM and Hamre BK (2008) *Classroom Assessment Scoring System: Prek Version*. Baltimore, MD: Brookes.

QRIS National Learning Network (2011) A foundation for quality improvement systems: state licensing, preschool, and QRIS program quality standards—an Office of Child Care issue brief. Available at: http://qrisnetwork.org/sites/all/files/resources/gscobb/2012-03-19%2012%3A52/Report.pdf

Shonkoff JP and Phillips DA (Eds) (2000) *From Neurons to Neighborhoods: The Science of Early Childhood Development*. Washington, DC: National Academy Press.

Tout K, Starr R, Soli M, et al. (2010) *Compendium of Quality Rating Systems and Evaluations*. Washington, DC: Office of Planning, Research, and Evaluation.

U.S. Department of Education (2013) Race to the top-early learning challenge. Available at: https://www.acf.hhs.gov/ecd/early-learning/race-to-the-top

U.S. Department of Education (2014) Preschool development grants. Available at: http://www2.ed.gov/programs/preschooldevelopmentgrants/index.html

Vandell DL (2004) Early child care: the known and the unknown. *Merrill-Palmer Quarterly* 50(3): 387–414.

Yoshikawa H, Weiland C, Brooks-Gunn J, et al. (2013) *Investing in Our Future: The Evidence Base on Preschool Education*. Washington, DC: Society for Research in Child Development & Foundation for child Development. Available at: https://www.fcd-us.org/assets/2016/04/Evidence-Base-on-Preschool-Education-FINAL.pdf

Zaslow M, Anderson R, Redd Z, et al. (2016) Quality thresholds, features, and dosage in early care and education: introduction and literature review. *Monographs of the Society for Research in Child Development* 81(2): 7–26.