**Abstract Title Page**

**Title:** Variation in Observed Program Characteristics across Classrooms in the Tennessee Voluntary Pre-Kindergarten Program

**Authors and Affiliations:**

Dale C. Farran
Peabody Research Institute, Vanderbilt University
dale.farran@vanderbilt.edu

Carol Bilbrey
Peabody Research Institute, Vanderbilt University
carol.bilbrey@vanderbilt.edu

## Abstract Body

### Background / Context:

Interest in scaling up prekindergarten programs for children from low income families is dramatically increasing, especially with the announcement by President Obama of support for "universal preschool" (Scott, 2013). A concern with expansion is that quality--and the ensuing effects–is hard to maintain when a program is taken to scale (Baker, 2011). Many states currently use the Early Childhood Environment Rating Scale – Revised edition (ECERS-R; Harms, Clifford, & Cryer, 2005) in combination with other measures to evaluate the quality of state-funded early care programs. In several states, both the ECERS-R and the Early Language and Literacy Classroom Observation (ELLCO; Smith & Dickinson, 2002) must be completed if a classroom serves preschool-age children. In addition to the general quality that is the focus of measures like the ECERS-R and ELLCO, classrooms differ widely in the time percentages allotted to different activities, with many classrooms spending an extremely large portion of the day in routine practices like bathroom breaks, meals, naptime, and transitions. Research in early education supports the idea of developmentally appropriate practice as reflected in instructional groupings and covered content (Bredekamp, 1987); however, it is not yet clear if such appropriate practice and consistency in that practice can be upheld in a statewide program. This paper focuses on evidence from one portion of the TN-VPK evaluation from which classroom observations yield information on classroom quality state-wide.

### Purpose / Objective / Research Question / Focus of Study:

This paper focuses on the variability in program characteristics associated with TN-VPK classrooms observed across the state of Tennessee.

### Setting:

As described in the previous papers' abstracts, TN-VPK supports 934 pre-k classrooms that serve more than 18,000 at-risk children in 133 of the 136 Tennessee school systems. It is, therefore, truly a statewide program serving all areas of the state, both rural and urban.

### Population / Participants / Subjects:

This paper includes a sample of 160 TN-VPK classrooms that were intentionally chosen to be representative of the distribution of TN-VPK program types across the state of Tennessee. Those classrooms were located in each of the state's four regions: 40 in West Tennessee (including Memphis), 39 in Central West Tennessee (including Nashville), 40 in Central East Tennessee (including Chattanooga), and 41 in East Tennessee (including Knoxville). This study is a partnership with the Tennessee Office of Early Learning (OEL) of the TN Department of Education. Administrators in OEL are responsible for the TN-VPK program. As the administrators in OEL requested, our sample of 160 included 43 classrooms that were part of the original pilot program that Tennessee began in 1998 and 117 classrooms that were put into place following that pilot period. The majority of the classrooms (118) were located inside of a public school, although the sample also included 35 classrooms off-site and 7 classrooms in locations

that served pre-k only. Per the requirements of the TN-VPK system in Tennessee, these 160 classrooms primarily served children from low-income families.

**Intervention / Program / Practice:**

A full description of the TN-VPK program is given in the abstract of the first paper in this symposium proposal.

**Research Design:**

The two previous papers have focused on the randomized control trial portion of the larger study of the effectiveness of the TN-VPK program. A second major component of the evaluation project involves an age-cutoff regression discontinuity design (RDD). Two cohorts of children in a representative sample of 160 classrooms across the state of Tennessee are participating in this phase of the study. Individual assessments on the same literacy, language, and math measures used in the RCT are being administered at the beginning of kindergarten for the cohort of children who were enrolled in TN-VPK in one of these schools during the previous school year. The scores for those children will then be compared with scores obtained from the cohort of children just entering TN-VPK in those same schools that same school year. With the difference in age accounted for statistically, this comparison will provide another set of estimates of the effects of TN-VPK. Although to date complete assessment data have been collected from only two of four state regions, the data from all four regions will provide a representative sample of TN-VPK programs statewide. These data will provide the state the opportunity to investigate whether programmatic characteristics are linked to differential child outcomes.

Unlike in the RCT portion of the evaluation, classrooms in the RDD portion are identified and all children in each classroom are assessed. Each classroom is observed for a full day. The OEL administrators identified "policy relevant" questions that they would like this portion of the full evaluation to address and observational measures were chosen accordingly. Eventually, the plan is to link observed differences among the classrooms to gains made by children in the classrooms. While we are not prepared to make those links in this paper, we are prepared to describe the differences found in program implementation, some of which are conceptually linked to the variation in effects obtained in the RCT sample of classrooms.

**Data Collection and Analysis:**

As part of the RDD, trained and reliable members of the PRI research team conducted in-depth systematic observations of the 160 pre-k classrooms. Using measures described below, we collected detailed information about the environments to which children who participate in TN-VPK are exposed. Observations took place on one day during the winter and lasted a minimum of four hours, covering the period of time prior to nap, typically considered the prime instructional time of the day in pre-k classrooms; 23 classrooms, spread evenly across the regions, were observed for the full day by two observers for inter-observer reliability. Measures included the following:

- Early Childhood Environment Rating Scale – Revised Edition (ECERS-R). The ECERS-R is one of the most widely-used instruments to evaluate the quality of early care and education environments. It is comprised of 43 items grouped into 7 subscales. Many

studies using the ECERS-R, including our study, use only the first six subscales (omitting the Parents and Staff subscale). The ECERS-R items are scored on a scale ranging from 1 (representing Inadequate quality) to 7 (Excellent quality). The total ECERS-R score is the simple average of the six subscale scores. ECERS reliability was scored as *exact* agreement between observers. With this strict criterion, observers achieved 88% agreement (Kappa = .84).

- Early Language and Literacy Classroom Observation (ELLCO). The ELLCO is an instrument used to evaluate the language and literacy environments of early childhood classrooms. The rating scale portion of the instrument involves 19 items grouped into 5 subscales (which are then further combined into 2 larger subscales and a total scale by averaging the small subscale scores). Each of the 19 items is rated on a 5-point scale with 1 representing Deficient quality and 5 representing Exemplary quality. In addition to these 19 items, the ELLCO includes a Literacy Environment Checklist which involves the ratings of presence/absence and quality of the physical environment in the classroom that pertains to language and literacy including books, writing materials, and classroom displays. ELLCO reliability was scored as *exact* agreement between observers. With this strict criterion, observers achieved 83% agreement (Kappa = .72).
- Narrative Record. The Narrative Record (Farran & Bilbrey, 2004) uses an open-ended format and coding scheme to record continuous data about the progression of activities in the classroom. It consists of observer notes and codes for each distinct episode that occurs, e.g., a small group activity. The data collected describe, at a minimum depending on which variable categories are included, the timing of the episode (start and stop times), the learning setting, and the learning focus. Because it is continuously recorded, the Narrative Record captures the entire observational period with no breaks in coding. The basic codes used in the Narrative Record across all four regions of the state covered primary divisions of activity setting used and instructional content of focus. Inter-observer agreement for the Narrative Record for 23 observation visits was defined by agreement within 3% of the total minutes observed in an episode type. For instance, if Observer 1 reported 30% of the observation spent in meals and Observer 2 reported 32% of that same observation spent in meals, the observer pair was deemed to be in agreement for that variable category. In reliability visits in participating classrooms, there was a mean observer agreement across the 11 Narrative Record activity setting categories of 94% (standard deviation of 7.3%). For content categories, mean observer agreement across the 10 categories was also 94% (standard deviation of 5.4%).

**Findings / Results:**

Table B1 presents descriptive statistics on summary scores for the three observation measures. Mean scores indicate the programs were, on average, less than what is considered "Good" quality according to ECERS-R anchors (i.e., score of "5"). On the ELLCO, mean scores all hovered near the anchor representing "Basic" evidence of quality. The Narrative Record showed that classrooms in this sample spent an average of 40% of the first four hours of the day, or 95 minutes, in routine activities including meals and transitions. Far more time was spent in whole-group activities (38%) than in any grouping involving center time (22%). What is more interesting than mean scores for the purposes of this paper, however, is the variation in scores across programs. Variation was present in all three classroom observation measures. For instance, within this scaled up prekindergarten program, there was one classroom with an

average ECERS score of 2.51, a score that the ECERS authors consider very close to Inadequate and therefore dangerous. Yet there was another program with an average score of 6.08, which is close to the maximum score on the scale and in the Excellent range. Similarly, in terms of how time was spent in the classroom, one classroom was observed spending almost 3 of its 4 hour morning observation in non-learning activities (routines, transitions, meals) while another spent 52 minutes of the 4 hours on those activities. Great variation was observed in how the day is structured and whether children are receiving most of their instruction in whole group teacher-led activities or spending some of their time in center-based, more independent activities.

**Conclusions:**

Our classroom observations of the statewide implementation of TN-VPK have demonstrated the same huge variability in program implementation that we saw in program effects in the RCT findings. Granted these are not the same classrooms and we have not yet examined the links between the observational variability obtained and child outcomes. However, at a larger level, we can conclude that statewide implementations of prekindergarten programs (for example, "Universal Pre-K") may have a difficult time ensuring quality control. The TN-VPK program is not alone in pushing to enact a program statewide without a clear vision communicated to its program participants of what should be happening in the classrooms. Moreover, as with other scaled up programs, few funds were allocated at a central level to supervise the quality of the individual classrooms or to determine and create the kind of professional development that may be needed. The children these programs are meant to serve are among the most vulnerable. Determining what aspects of pre-k programs should be uniform in order to serve the children well is an important, overdue step.

# Appendices

## Appendix A. References


Baker, M. (2011). Innis Lecture: Universal early childhood interventions: what is the evidence base? *Canadian Journal of Economics/Revue canadienne d'economique, 44* (4), 1069-1105.

Bredekamp, S. (1987). *Developmentally Appropriate Practice in Early Childhood Programs Serving Children from Birth through Age 8.* Washington, D.C.: National Association for the Education of Young Children.

Farran, D. C., & Bilbrey, C. (2004). *Narrative Record.* Unpublished instrument available from D.C. Farran, Peabody Research Institute, Vanderbilt University, Nashville, TN.

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

Smith, M. W., & Dickinson, D. K. (2002). *User's guide to the Early Language and Literacy Classroom Observation toolkit.* Baltimore, MD: Brookes.

## Appendix B. Tables and Figures

Table B1

*Descriptive Statistics for Observation Measures – Subscales and Total Scores*

| | Minimum | Maximum | Mean | SD |
|---|---|---|---|---|
| ECERS | | | | |
| Space and Furnishings | 2.63 | 6.13 | 4.23 | 0.66 |
| Personal Care Routines | 1.80 | 6.50 | 3.79 | 1.02 |
| Language and Reasoning | 1.75 | 7.00 | 4.64 | 1.08 |
| Activities | 1.56 | 6.30 | 3.15 | 0.86 |
| Interactions | 2.00 | 7.00 | 5.39 | 1.05 |
| Program Structure | 1.33 | 7.00 | 3.99 | 1.22 |
| Total Score | 2.51 | 6.08 | 4.20 | 0.75 |
| ELLCO | | | | |
| Classroom Structure | 2.25 | 5.00 | 3.81 | 0.59 |
| Curriculum | 1.67 | 4.67 | 3.26 | 0.62 |
| Language Environment | 2.00 | 5.00 | 3.37 | 0.65 |
| Books and Book Reading | 1.60 | 5.00 | 3.46 | 0.58 |
| Print and Early Writing | 1.33 | 5.00 | 3.14 | 0.70 |
| Total Score | 1.88 | 4.79 | 3.43 | 0.50 |
| Narrative Record | | | | |
| Activity Setting | | | | |
| % Time in Routines | 0.22 | 0.68 | 0.40 | 0.08 |
| % Time in Whole-Group Instruction | 0.10 | 0.68 | 0.38 | 0.11 |
| % Time in Instruction Involving Centers | 0.00 | 0.52 | 0.22 | 0.10 |
| Content of Focus | | | | |
| % Time Focused on Literacy | 0.00 | 0.48 | 0.16 | 0.08 |
| % Time Focused on Math or Science | 0.00 | 0.20 | 0.04 | 0.05 |
| % Time Focused on Mixed Content | 0.04 | 0.56 | 0.29 | 0.11 |
| % Time Without Any Instructional Content | 0.20 | 0.62 | 0.38 | 0.08 |