

Title: Measuring Preschool Program Quality: Multiple Aspects of the Validity of Two Widely-Used Measures

Topic/Goal: Early Learning Programs and Policies, Goal One: Exploration

Purpose: The validity of measures of preschool quality is a pressing policy concern, as state Quality Rating and Information Systems, Head Start recompetition regulations, and the Race to the Top Early Learning Challenge rely on these measures. Our proposed study will inform policy and research by examining understudied aspects of validity of two widely used measures: the Early Childhood Environment System Rating Scale – Revised (ECERS-R) which is used in most states' QRIS and the Classroom Assessment Scoring System (CLASS) which is used to monitor Head Start programs and two states' QRIS.

We propose to examine predictive validity specific to the cutoffs defined in policy systems, translating significant associations into effect sizes for researchers and proportions of school readiness gaps for policymakers. Particularly important for policymakers is our use of a dozen datasets with the ECERS-R measure, allowing us to look for the replication essential for high stakes decision making. We also propose to examine structural validity, testing for the dimensionality of each scale. Prior factor analyses of the ECERS-R have not always been published nor publicized to policymakers, and state QRISs rely on the ECERS-R total score, pointing to the importance of verifying the ECERS-R dimensions across datasets and making the results accessible to decision makers. We will also use item response theory models to test for order (whether higher scores reflect higher quality on the underlying dimension), fit (whether items go together to define an underlying dimension) and separation (the degree to which items can distinguish low, moderate and high quality centers). Only one other published study of U.S. preschools has evaluated the ECERS-R or the CLASS with these approaches (Gordon, Fujimoto, et al., 2012, a product of our recently completed IES grant). To the extent there is modest predictive and structural validity of the measures, our IRT findings about item quality may help explain why and point to directions for future measure development.

Setting: The project will use twelve datasets, most drawn with national sampling frames or from multiple sites across the country. In each data set, we select only the 3-5 year old children who receive care or attend preschool in centers. Depending on the dataset, these programs are located in public and private schools, churches, and non-profit or for-profit organizations.

Population: Some studies focus on at-risk subpopulations, including low-income children, Head Start participants, and state pre-kindergarten programs. Others are nationally representative but include a large number of low-income children.

Primary Research Method: Our proposed research relies on secondary data analyses and meta-analyses.

Key Measures and Key Outcomes: We measure preschool quality with the ECERS-R and CLASS. Child cognitive outcomes include the Peabody Picture Vocabulary Test and Woodcock-Johnson. Child socio-emotional outcomes include the Social Skills Rating System and Teacher Report Form. Child health outcomes include the child's overall health and body mass index.

Data Analytic Strategy: We conduct factor analyses to examine structural validity, regression analyses to examine predictive validity, and item response theory analyses to examine response process and test content validity. Regression analyses examine non-linear associations with dummy variable and piecewise linear models. We meta-analyze regression coefficients and interpret significant regression coefficients in terms of effect sizes and portion of school readiness gaps (with gaps estimated by national data or meta-analysis of existing literature).