

EDUCATION

# Can Rating Pre-K Programs Predict Children’s Learning?

T. J. Sabol,<sup>1\*</sup> S. L. Soliday Hong,<sup>2</sup> R. C. Pianta,<sup>3</sup> M. R. Burchinal<sup>2</sup>

Early childhood education programs [e.g., prekindergarten (pre-K)]—characterized by stimulating and supportive teacher-child interactions in enriched classroom settings—promote children’s learning and school readiness (1–3). But in the United States, most children, particularly those from low-income backgrounds, attend programs that may not be of sufficient quality to improve readiness for school success (4). States are adopting Quality Rating and Improvement Systems (QRISs) as a market-based approach for improving early education, but few states have evaluated the extent to which their QRIS relates to child outcomes. We studied the ability of several QRISs to distinguish among meaningful differences in quality that support learning.

QRISs are designed to rate individual programs and to disseminate ratings to inform parents’ decisions about enrollment and

policy-makers’ decisions about investment, so as to create a local market for quality improvement intended to lead to better school readiness (5). Because of increased federal and state support, over half of all U.S. states now have a QRIS (5). The U.S. Department of Education invested \$500 million in selected states in the Race to the Top—Early Learning Challenge program to expand QRISs.

Development of each state’s QRIS is typically based on local professionals’ judgment and involves (i) selecting measures of “quality” that form the individual indicators; (ii) categorizing each indicator into different levels of quality, and (iii) creating rules for how to combine indicators into composite ratings. Differing from many accountability policies, QRISs typically do not measure child outcomes to assess school quality. Instead, they assess process-oriented and structural indicators (e.g., teachers’ qualifications). Some indicators, such as quality of the learning environment, are modestly related to children’s skills (2). There is less evidence that other indicators, such as family partnership, have a robust relation to child outcomes (6). Small-scale QRIS evaluations have shown mixed evidence as to whether ratings discriminate among pro-

Systems intended to rate the quality of early education programs may not actually reflect impacts on learning.

grams that are better or worse in fostering learning (7). Neither the selection of quality indicators, their categorization into levels, nor the composite ratings have been systematically studied as they relate to children’s learning.

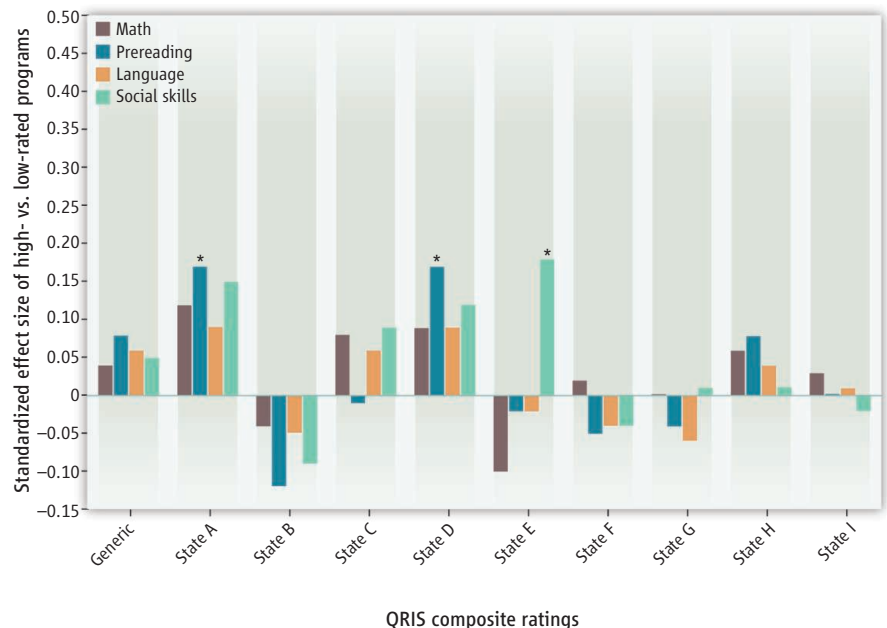
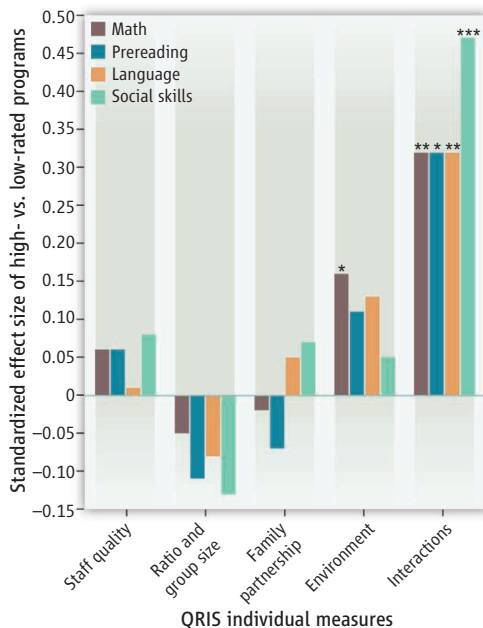
### Replicating Systems, Relating Outcomes

We estimated associations between QRIS ratings and measures of children’s learning using a data set that included 2419 children in 673 public pre-K programs across the United States [see supplementary materials (SM) for details]. Data are from two studies that represented the variety of pre-K models in the United States at the time (2001–04): the National Center for Early Development and Learning (NCELD) multistate study of prekindergarten and the State-Wide Early Education Programs (SWEEP) Study. The studies used a stratified random sample

**On most measures of children’s learning, programs rated high by QRIs produce outcomes that are not significantly better than those of low-rated programs.** Stars indicate a statistically significant difference in math, prereading, expressive language, and social skills (\* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ ), see SM. (Left) QRIS individual measures. (Right) Composite ratings, see (10).

<sup>1</sup>Institute for Policy Research, Northwestern University, Evanston, IL 60208, USA. <sup>2</sup>Frank Porter Graham Child Development Institute, University of North Carolina, Chapel Hill, NC 27599, USA. <sup>3</sup>Center for the Advanced Study of Teaching and Learning, University of Virginia, Charlottesville, VA 22904, USA.

\*Corresponding author: terri.sabol@northwestern.edu



design to select programs within each state system (8).

We used the NCEDL-SWEEP data set to replicate as closely as possible a number of individual state's QRIS scoring algorithms (see SM). These QRISs used the same four popular quality indicators to determine program ratings: (i) staff qualifications, including teacher and director level of education and years of experience; (ii) staff-child ratio and group size; (iii) family partnerships; and (iv) learning environment, measured by the Early Childhood Education Rating System-Revised (ECERS-R) (9, 10).

We also selected an additional indicator, (v) interactions, measured using the Classroom Assessment Scoring System (CLASS) (11) to assess the quality of interactions in the classroom. None of the selected existing QRISs used the CLASS in program ratings. We created and tested a generic system that included all five quality indicators [(i) to (v) above], representing the most simplified and best predictor scoring system we could design based on available data (see SM).

All analyses examined the relation among QRIS measures and end-of-year child outcomes, accounting for nesting in classrooms; beginning-of-year scores; state location of the pre-K program; and child, family, and program characteristics. All programs in the sample were subject to states' minimum requirements in terms of structural features and teacher qualifications. Our findings can be interpreted as an estimate of the association among quality measures and child outcomes beyond these minimum requirements.

#### Few Associations of QRISs with Outcomes

Of the five individual quality indicators, the CLASS measure of teacher-child interaction quality consistently was the strongest predictor of children's learning, followed by the learning environment (ECERS-R; see SM). The structural quality measures of staff qualifications, staff-child ratio, and family partnership provided weaker and less consistent prediction of children's learning.

The relation of child outcomes when children attended a program rated in the highest level versus the lowest level for each quality indicator is shown in the chart (left) and is based on cut-points from our generic QRIS. Results were similar when we used cut-points for the nine states regardless of small variations in scoring methods (see SM). The high-low contrast on the measure of teacher-child interactions had the strongest association with children's skills. Children had stronger math skills, but not prereading, expressive language, or social

skills, when they attended programs with the highest environmental rating scale compared with the lowest rating scale.

We then aggregated individual quality indicators according to generic and state QRIS scoring rubrics, converted to overall composite ratings, and compared the relation among high- and low-rated programs on child outcomes (see the chart, right). Generic system composite ratings were not significant predictors of child outcomes. There were few associations of composite ratings to children's learning in the nine replicated state QRISs. The association among composite ratings and child outcomes were about half the size of that for the single indicator of teacher-child interactions.

There are important limitations of our approach. Most state QRISs are designed to be used for child care and early-learning programs serving children from birth through preschool ages and typically incorporate both public and private programs. The NCEDL-SWEEP studies only include public pre-K programs serving preschool-age children. We model the association between QRIS ratings and child outcomes by controlling for beginning-of-year scores that may reflect preexisting differences between children. This value-added approach has limitations that may lead to biased estimates (12, 13). We do not show evidence of causal impacts of QRIS indicators and composites on children's learning.

#### Implications for QRISs

For QRISs to increase the likelihood of pre-K programs improving children's school readiness, programs that receive the highest composite ratings should improve children's skills more than programs that receive the lowest ratings. As presently designed and implemented, QRIS evaluation may not realize these aims. Although individual observational measures of quality, particularly teacher-child interactions, were related to children's learning, once multiple indicators were aggregated and converted to ratings, they failed to yield consistent relations to learning.

States should capitalize on the rich information they are collecting and consider new methods for analyzing and disseminating it. Given costs associated with collecting multiple indicators of quality, states may conserve resources and enhance the validity of QRIS ratings by focusing on indicators with demonstrable links to children's learning. To provide easily interpretable levels of quality to parents and communities, states often transform levels of measurements from continuous to ordinal scales. Some states already present information for each individual qual-

ity indicator, without aggregating multiple indicators to form one composite rating.

These findings may have implications for other education accountability efforts that attempt to translate ratings to the public but, in so doing, convert multiple assessments of performance that are only obliquely related to learning, which obscures possibly important effects of key indicators.

Observational measures of teacher-child interactions demonstrate promise for discerning meaningful levels of pre-K quality. Yet QRISs are not capitalizing on this empirical evidence. As QRISs become more commonplace and pre-K programs plan to expand enrollments, it is increasingly important that ratings link to children's learning to ensure that states incentivize and improve the aspects of quality that matter most.

#### References and Notes

1. W. S. Barnett, *Science* **333**, 975–978 (2011).
2. R. C. Pianta, W. S. Barnett, M. R. Burchinal, K. R. Thornburg, *Psychol. Sci.* **10**, 49–88 (2009).
3. W. T. Gormley Jr., D. Phillips, T. Gayer, *Science* **320**, 1723–1724 (2008).
4. J. Jacobson-Chernoff, K. Flanagan, C. McPhee, J. Park, *Preschool: First Findings From the Preschool Follow-up of the Early Childhood Longitudinal Study, Birth Cohort (ECLS-B)* (National Center for Education Statistics, U.S. Department of Education, Washington, DC, 2007).
5. K. Tout et al., *The Child Care Quality Rating System Assessment: Compendium of Quality Rating Systems and Evaluations* (Office of Planning, Research & Evaluation, Administration for Children & Families, U.S. Department of Health and Human Services, Washington, DC, 2010).
6. G. Camilli et al., *Teach. Coll. Rec.* **112**, 579–620 (2010).
7. G. Zellman, M. Perlman, M., V. Le, C. M. Setodji, *Assessing the Validity of the Qualistar Early Learning QRIS as a Tool for Improving Child-Care Quality* (RAND Corporation, Santa Monica, CA, 2008).
8. D. Early et al., "Pre-kindergarten in eleven states: NCEDL's Multi-State Study of Prekindergarten and State-Wide Early Education Program Study" (Tech. Rep. Frank Porter Graham Child Development Institute, Chapel Hill, NC, 2005).
9. T. Harms, R. M. Clifford, D. Cryer, *The Early Childhood Environment Rating Scale: Revised Edition* (Teachers College Press, New York, 1998).
10. Because we were using a common data set to replicate individual QRISs, we were not able to include all quality indicators used by a state, and in the case of family partnership or staff qualifications, our measures often did not align exactly with states' measures (see SM). For this reason, we removed identifying names of each state in the right side of the figure.
11. R. C. Pianta, K. La Paro, B. K. Hamre, *Classroom Assessment Scoring System Manual: Pre-Kindergarten* (Brookes, Baltimore, MD, 2004).
12. S. W. Raudenbush, *J. Educ. Behav. Stat.* **29**, 121–129 (2004).
13. J. Rothstein, *Educ. Finance Policy* **4**, 537–571 (2009).

**Acknowledgments:** The authors are supported by Institute of Education Sciences, U.S. Department of Education, Grant R305A060021. R.C.P. is coauthor of tCLASS and cofounder and shareholder of Teachstone Training, LLC, a company focused on measuring and improving teacher-student interactions. The University of Virginia owns an equity share in Teachstone.

#### Supplementary Materials

[www.sciencemag.org/cgi/content/full/341/6148/845/DC1](http://www.sciencemag.org/cgi/content/full/341/6148/845/DC1)

10.1126/science.1233517