

# An Inquiry into Pennsylvania's Keystone STARS: SUMMARY REPORT

Phil Sirinides, PhD John Fantuzzo, PhD Whitney LeBoeuf, PhD Katie Barghaus, PhD Ryan Fink, EdD

#### **NOVEMBER 2015**

The full report is available on the University of Pennsylvania's Consortium for Policy Research in Education website at cpre.org.

This study was conducted with financial support from the William Penn Foundation and prepared for Pennsylvania's Office of Child Development and Early Learning (OCDEL). The opinions expressed in this report are those of the authors and do not necessarily reflect the views of the William Penn Foundation or OCDEL.

**H** igh quality care in the earliest years of life has been shown to relate to positive developmental outcomes for children, including improved early academic skills, social-emotional competencies, and cognitive functioning.<sup>1</sup> Unfortunately, the early care experiences of many children are not always high quality; rather, research suggests that high-quality care is the exception.<sup>2</sup> The growing evidence relating quality care to improved outcomes, the variability in quality across care settings, and the failure of existing approaches to improve child care have led to a national call to enhance the quality of early care and education programs.<sup>3</sup> In response to this call, states have created *Quality Rating and Improvement Systems (QRISs)*.

The ultimate goal of a state QRIS is to assist service providers in the delivery of quality early care and education in order to improve children's developmental outcomes.<sup>4</sup> Fundamentally, all QRISs include: (1) an emphasis on improved child outcomes, (2) quality components, which are sets of related performance standards for early care and education expected to influence child outcomes, and (3) a system reflecting a tiered approach to measuring provider quality and guiding improvements. Since their inception almost 15 years ago, QRISs have been implemented in 39 states either statewide or locally.

Pennsylvania's QRIS, Keystone STARS, was one of the first systems in the nation. Launched statewide in 2003, the system consists of 12 quality components: (1) Director Qualifications, (2) Director Development, (3) Staff Qualifications, (4) Staff Development, (5) Child Observation, Curriculum and Assessment, (6) Environment Rating, (7) Community Resources and Family Involvement, (8) Transition, (9) Business Practices, (10) Continuous Quality Improvement, (11) Staff Communication and Support, and (12) Employee Compensation.<sup>5</sup> Child care and Head Start providers that voluntarily participate in Keystone STARS must meet all performance standards at each of the system's four STAR levels before receiving the corresponding quality rating.<sup>6</sup> A rating of STAR 1 is considered the lowest quality level and a rating of STAR 4 is considered the highest level.

## INQUIRY OBJECTIVES

A team from the University of Pennsylvania was funded by the William Penn Foundation to conduct an inquiry of Keystone STARS. The goal of this inquiry was to provide a broad look at Keystone STARS to inform future revisions and evaluation of the system as part of Pennsylvania's Race to the Top Early Learning Challenge grant (2013-2018). The inquiry focused on providing an overarching look at Keystone STARS with respect to three major areas:

- 1. Child outcomes. This inquiry examined the relations between Keystone STARS and children's overall developmental competencies.
- 2. Quality components. This inquiry investigated the extent of evidence from theory, empirical research, and practitioner expertise linking each of the Keystone STARS quality components to child outcomes.
- 3. Systems approach to rating quality and guiding improvements. This inquiry examined overall features of the system that could be improved to enhance the effectiveness and efficiency of the system.

## CHILD OUTCOMES

### Data

This inquiry investigated the relationship between Keystone STARS levels (e.g., STAR 1, STAR 2, etc.) and children's developmental outcomes, as well as the relationship between Keystone STARS quality components (e.g., Staff Qualifications, Transitions, etc.) and children's developmental outcomes. Outcome data were obtained in Spring 2015 using the Work Sampling System (WSS) for a sample of 1,108 four-year-olds from all five regions of Pennsylvania.<sup>7</sup> Only a WSS total score was used in this study. Data came from 11 STAR 1 centers, 9 STAR 2 centers, 15 STAR 3 centers, and 14 STAR 4 centers.

### **Findings**

The majority of children received a high total score on the WSS (75% of children scored above 2.5 on the three-point scale for the total score). Therefore, the inquiry team compared the median outcome scores across STAR levels and tested group differences using an approach that accounts for the fact that high scores were more common than medium and low scores.<sup>8</sup> Figure 1 shows median scores and 95 percent confidence intervals for each STAR level.

- ▶ Four-year-old children in STAR 3 and 4 centers performed significantly higher on the WSS total score than those in STAR 1 and STAR 2 centers, though estimated effects were small.
- No difference in WSS total scores was found between STAR 1 and 2 centers.
- No difference in WSS total scores was found between STAR 3 and STAR 4 centers.

Only the Environment Rating quality component had sufficient evidence-based measurement to empirically explore its relation to children's outcomes.<sup>9</sup> This component uses the Early Childhood Environment Rating Scale-Revised (ECERS-R). Correlations were used to examine the associations between the WSS total score and the ECERS-R total and subscale scores.<sup>10</sup>

• Environment quality ratings, as measured by ECERS-R, were positively and statistically significantly associated with WSS total scores, although these estimates were small.

#### Figure 1. Average WSS Score by STARS Level



- Scores on three of the seven ECERS-R subscales (Space and Furnishings, Activities, and Program Structure) were found to be positively associated with WSS total scores.
- Correlation coefficients between the WSS total scores and ECERS-R subscales of Personal Care Routines, Language-Reasoning, Interactions, and Parents and Staff were all non-significant.

### QUALITY COMPONENTS Data

To investigate the extent of evidence currently available for each of the STARS quality components, the research team examined three different sources of data:

- Child development theory. The inquiry team used the developmentalecological model to determine the theoretical level of influence of each quality component in the Keystone STARS system on child development.<sup>11</sup> This model of human development served as the basis for federal and state standards for early childhood care and education. Center-based performance standards for each of the STARS quality components were reviewed to understand how the components were defined in the system. Based on how these quality components were operationalized for centers, the research team sorted the components by their theoretical level of influence on child development.
- ▶ Existing empirical research. The research team performed a systematic search for research on the relationships between quality components in QRISs and child outcomes. The team intentionally focused on studies performed within the context of a QRIS in order to understand how each quality component, as defined and operationalized through these systems, may relate to child outcomes. Only six studies explicitly evaluated the relationship between QRIS components and child outcomes.<sup>12</sup> For each of the STARS quality components, the inquiry team documented the number of: (1) studies that examined its relationship to child outcomes, (2) significant results in the expected direction, (3) significant results in the unexpected direction, and (4) tested relationships that were not significant.
- ▶ Keystone STARS provider experiences with quality components. The inquiry team administered a survey that asked providers to identify components of quality they believed to be related to child outcomes.<sup>13</sup> Quality components ranked in the top third of all components in terms of importance were categorized as having high importance for child outcomes. Components ranked in the bottom two-thirds of all components were categorized as having moderate to low importance for outcomes. All components that were grouped in the top third were statistically significantly different than all components in the bottom third.

### **Findings**

The inquiry team synthesized the data from these three sources of evidence and visually summarized the findings in the figure below. This figure represents the amount of evidence supporting each quality component's direct relationship to child outcomes. Components which currently have the most evidence are situated in the inner circle, while those with less appear in the outer circles.

#### Figure 2. Inquiry-based Organization of Quality Components



- Supporting individual child learning. The innermost circle includes the two quality components with multiple sources of evidence: Child Observation, Curriculum, and Assessment and Environment Rating. Using the developmental-ecological model, these quality components were found to most closely support individual child development. Providers indicated that the Child Observation, Curriculum, and Assessment component was highly important for improving child outcomes. Some empirical evidence was found to support the connection between Environment Rating and child outcomes. These quality components represent a common goal of directly "supporting individual child learning."
- Strengthening teacher and family interactions with children. The middle circle represents quality components with one source of evidence linking them to child outcomes: Transition, Staff Qualifications, Staff Development, Community Resources and Family Involvement, and Staff Communication and Support. As noted in the figure, these five quality components serve the common goal of "strengthening teacher and family interactions with children."

Sustaining the child care provider. The outermost circle includes the five quality components for which none of the evidence sources examined linked them directly to child outcomes: Director Development, Director Qualifications, Employee Compensation, Continuous Quality Improvement, and Business Practices. It is logical that these quality components do not have any clear evidence directly linking them to child outcomes because they are designed to "sustain the child care provider." These components are important for the overall sustainability and success of a child care and education setting. The potential influence of these components on children's development and learning is indirect. These components encourage providers to establish stable, sustainable businesses, which in turn may help to create a more positive educational climate for children.

### SYSTEMS APPROACH TO RATING QUALITY AND GUIDING IMPROVEMENTS Data

For the systems investigation, the research team examined two different data sources:

- Perspectives of Keystone STARS Developers and System Administrators. Interviews were conducted with 14 developers and/or implementers of Keystone STARS.<sup>14</sup> The interviews were guided by a semi-structured interview protocol exploring: the respondent's role in Keystone STARS; the origin of quality components and standards; perception of providers' experiences with the system; and the evolution of Keystone STARS.
- Perspectives of Keystone STARS Providers. The survey of providers asked questions about their experiences with Keystone STARS, including their reasons for participating in the program, motivation for moving up in the system, and challenges to meeting particular standards. Providers were also given an opportunity to share their perspectives about Keystone STARS through open-ended questions. These data contributed a provider perspective to guide and enhance system improvements.

### **Findings**

The investigation analyzed data from developers, system-level implementers, and providers to assess how the STARS system functioned from their perspective. This examination revealed three system challenges:

- ▶ Too many standards unrelated to child outcomes. System-level program administrators and child care providers both expressed a belief that Keystone STARS currently has too many requirements and that many requirements are not directly related to improved child outcomes. They indicated that there are system requirements that divert attention and resources away from the primary goal of preparing children for school.
- Requirements are overly prescriptive. Motivating and incentivizing providers to remain engaged in a quality improvement process has been a challenge for STARS program administrators. Providers, for their part, view the system largely as one of compliance.

Inconsistent progression of expectations across STAR levels. Although Keystone STARS was intended to be a roadmap to quality, some providers experience the transition between levels as disjointed and feel stuck at their current level of quality.

## LESSONS LEARNED

Findings from this inquiry produced several key lessons, which may influence future work examining Keystone STARS and other QRISs:

- ▶ High quality, measurable indicators of child outcomes and quality components are needed. Child outcome data currently reported is insufficient to assess the relationships of STAR levels and STAR components to child outcomes. This highlights the need for more sensitive measures of children's developmental outcomes. In addition, only the Environment Rating quality component had sufficient data to examine its relationship to child outcomes. This discovery indicates a need for measurement of the other 11 quality components so future efforts can assess their relationships to child outcomes.
- Research on associations among child outcomes and quality components is needed. The empirical QRIS research base consists of a limited number of studies examining the relationships between quality components and child outcomes. This research is characterized by predominantly non-significant findings and lacks consistency across studies when findings are significant. As a whole, this makes drawing broad conclusions about the importance of specific components for positive child outcomes difficult. More research on the components hypothesized to have the most direct and substantial influence on child outcomes within the QRIS setting is needed, and QRISs must evolve as new information is generated.
- The overarching logic and purpose of STARS should be revisited. As revisions to Keystone STARS are now being considered, it is critical that its overall logic and purpose is reexamined in collaboration with providers and other stakeholders. Ensuring consensus on these primary points will provide a road map for refinements to the system.

## RECOMMENDATIONS

**Make relevant distinctions among system requirements to prioritize improving child outcomes.** While many quality components and standards were initially included in the system to comprehensively improve child care settings, it is time to prioritize requirements that demonstrate the greatest value for improving developmental outcomes for young children in Pennsylvania. This recommendation is supported by QRIS research which calls for focusing on the "few and powerful" quality components with demonstrable links to child outcomes.<sup>15</sup> The creation of three program tracks (illustrated below) represents a possible method of streamlining the system to account for these distinctions.



Monitoring and Reporting. Like all public programs, STARS needs capacities for its own monitoring and improvement. This track is primarily intended to maintain integrity and efficiency in program operations, support system-level quality improvement, and generate evidence of the programs' outcomes for funding and sustainability.

**Define Keystone STARS as steps to quality and not levels of quality.** The original intention of system developers was to have Keystone STAR levels serve as steps to quality and not necessarily levels of quality. It is important to reclaim this feature of the system. After STARS requirements have been streamlined, the progression of expectations across STAR levels should be clearly specified within each of the tracks outlined above. A meaningful reorganization of standards will help providers understand the progression of expectations across STAR levels for each track.

- For the evidence-based standards track. STAR 1 providers complete all preparation necessary to begin quality improvement activities. By STAR 2, providers engage in improvement activities that lead to meeting the evidence-based definition of quality. By STAR 3, providers are deeply engaged in improvement activities with demonstrable progress toward meeting quality. By STAR 4, valid and reliable measurement indicates that providers have met evidence-based performance standards.
- For the individual improvement activities track. The Plan, Do, Study, Act progression could be implemented to accommodate the progression of individualized goals.<sup>16</sup> At STAR 1, providers establish an action plan with performance metrics (Plan). At STAR 2, providers implement elements of the action plan (Do). By STAR 3, providers record performance metrics to learn about challenges, opportunities, and achievements, gaining input from a range of data sources and stakeholders (Study). Finally, by STAR 4, providers design and implement changes to address challenges and opportunities for improvement (Act).

#### **Figure 3. Tracks for Program Requirements**

• For monitoring and reporting. Expectations would be placed at each STAR level as needed, such that they serve the needs of system improvement while not overburdening providers.

**Create a Logic Model to Guide Revisions.** In order to pursue these next steps and revise Keystone STARS based on the lessons learned from this inquiry, Pennsylvania needs to develop a road map, or logic model, to guide revisions and system operations going forward. A logic model is a systematic and visual way to present expected causal links among inputs, activities, and outputs and desired outcomes.<sup>17</sup> Well-developed logic models can be used: as a road map for system changes and operations, to identify where measurement is needed to monitor provider progress, and as a tool that can communicate how expectations relate to overall system goals. There is national recognition of the importance of logic models to the success of QRISs; however, only eight states have publicly available models specifically detailing the operations of their QRIS.<sup>18</sup> Pennsylvania has an opportunity to advance the field by developing a comprehensive logic model.

## ENDNOTES

1 Burchinal, Kainz, Cai, Tout, Zaslow, Martinez-Beck, & Rathgeb, 2009; National Institute of Child Health and Human Development Early Child Care Research Network, 2000, 2005; Vandell, 2004.

2 Fiene, Greenberg, Bergsten, Fegley, Carl, & Gibbons, 2002; Karoly, Ghosh-Dastidar, Zellman, Perlman, & Fernyhough, 2008.

3 Karoly, Zellman, & Perlman, 2013

4 Zellman, Perlman, Le, & Setodji, 2008

5 For family child care home and group home providers quality components that relate to Director and Staff are identified as Primary Staff Person and Secondary Staff Person.

6 There were two pathways by which a program could be ranked at the STAR 4 level: (1) by meeting all of the performance standards for level 4, or (2) by demonstrating current accreditation from an OCDEL-accepted program and provide evidence that a specific subset of STARS standards have been met. Programs rated at STAR level 4 by these two pathways were analyzed separately. Results are presented here for the 14 program that were ranked at the STAR 4 level by meeting all of the performance standards for level 4 (i.e., pathway 1). Results for the four centers that meet the STAR 4 level through accreditation and providing evidence that they had meet a specific subset of STARS standards are not reported here (please see the full report for these results).

7 Preliminary analysis showed insufficient psychometric support for using the WSS subscale scores. In addition, data were also collected on a geographically diverse sample of 1,142 three-year-olds. However, insufficient concurrent validity evidence was found to support the use of outcomes from three-year-olds. Thus, findings from this study's data only provided support for using the WSS Total Score for 4-year-olds in subsequent analyses.

8 Differences between median scores rather than mean scores were used because this approach is not influenced by the skewness of Spring WSS scores. WSS total scores range from 1 to 3. Group differences were tested using non-parametric bootstrapped standard errors.

9 Criteria for determining if each quality component had sufficient data included: (1) reliable measurement of quality that was able to detect variation across centers withi STAR levels and (2) independence from the performance standards such that the data indicated something about the degree of quality and not simply whether or not standards were met.

10 Pearson and Spearman rank correlation coefficients were used to examine the associations between the WSS total score and the ECERS-R total and subscale scores.

11 Bronfenbrenner, 1994

12 Elicker, Langill, Ruprecht, Lewsader, & Anderson, 2011; Hestenes, Kintner-Duffy, Wang, La Paro, Mims, Crosby, Scott-Little, & Cassidy, 2014; Peisner-Feinberg, LaForrett, Schaaf, Hildebrandt, Sideris, & Pan, 2014; Sabol, Hong, Pianta, & Burchinal, 2013; Tout, Starr, Isner, Cleveland, Albertson-Junkans, Soli, & Quinn, 2011; Zellman, Perlman, Le, & Setodji, 2008

13 The survey sample was drawn from the population of all child care providers who were participating in Keystone STARS as of summer 2014. Responses were submitted by 672 providers (70% response rate of active providers) representing all provider types and STAR levels.

14 Four of the individuals were independent from both OCDEL and state contractors affiliated with Keystone STARS. The remaining ten interviewees were either former or current employees of OCDEL or a contractor.

15 Stoney, 2014; Yoshikawa et al., 2013

16 The Plan, Do, Study, Act Cycle is a quality improvement approach that has been adapted and applied in a number of fields since it was first introduced by W. Edwards Deming in his 1986 book, Out of the Crisis.

17 Lugo-Gil, Sattar, Ross, Boller, Kirby, & Tout, 2011

18 The research team systematically searched for state QRIS logic models and only located eight models.