RESEARCH REPORT

Evaluation of Delaware Stars for Early Success

Final Report

Lynn A. Karoly, Heather L. Schwartz, Claude Messan Setodji, Ann C. Haas

Sponsored by the Delaware Office of Early Learning

RAND
CORPORATION

For more information on this publication, visit www.rand.org/t/rr1426

Library of Congress Cataloging-in-Publication Data is available for this publication.

ISBN: 978-0-8330-9666-1

Published by the RAND Corporation, Santa Monica, Calif.

Preface

In June 2013, the Delaware Office of Early Learning contracted with the RAND Corporation to conduct an independent evaluation of Delaware Stars for Early Success, the state's quality rating and improvement system (QRIS) for early learning and care programs. The purpose of the RAND evaluation was to support Delaware in its efforts to design and implement an effective, robust system for measuring and reporting on the quality of early learning and care programs in home and center settings. The evaluation further aimed to inform efforts to improve the quality of programs in ways that are beneficial for participating children and their families. The project entailed a series of interrelated research tasks designed to provide objective and rigorous empirical evidence of the extent to which rating tiers reflect relevant differences in the quality of home- and center-based programs and whether the system is operating effectively in terms of technical assistance (TA), financial support, and other features.

In support of the overall evaluation goals, this final report summarizes the findings from all components of the evaluation. Our primary focus is on addressing two sets of questions central to the evaluation:

- Do early care and education (ECE) programs with higher ratings in the QRIS deliver higher-quality care and early learning than those with lower ratings? What is the relationship between program characteristics and quality in Delaware Stars?
- Do children in programs with higher ratings in the QRIS have better learning and developmental outcomes than children in programs with lower ratings? What dimensions of Delaware Stars program ratings are most vital to child learning and developmental outcomes?

These questions are examined using data collected in 2014–2015 on program quality from a sample of Delaware ECE providers, along with measures of learning for children enrolled in the sampled programs. We also report on results from a survey of the directors of the sampled providers. Other components of the evaluation are also addressed in this report, including findings regarding quality improvement supports, financial incentives, and other aspects of system performance, some of which were analyzed in more detail in two earlier reports.

The findings of this third and final report will be of interest to stakeholders in Delaware, as well as those in other states who are involved in designing, implementing, or evaluating ECE QRISs. The findings from the first two reports from the evaluation are found in

- Heather L. Schwartz, Lynn A. Karoly, Vi-Nhuan Le, Jennifer Tamargo, and Claude Messan Setodji, *Evaluation of Delaware Stars for Early Success: Year 1 Report*, Santa Monica, Calif.: RAND Corporation, RR-606-DOEL, 2014 (www.rand.org/t/RR606)

- Anamarie Auger, Lynn A. Karoly, and Heather L. Schwartz, *Evaluation of Delaware Stars for Early Success: Year 2 Report*, Santa Monica, Calif.: RAND Corporation, RR-1026-DOEL, 2015 (www.rand.org/t/RR1026).

This research was conducted jointly in RAND Education and RAND Labor and Population. Additional information about RAND is available at www.rand.org.

Table of Contents

Figures

Tables

Summary

As of May 2016, Delaware is one of 39 states with a fully implemented statewide early care and education (ECE) quality rating and improvement system (QRIS). States and localities have been designing, piloting, and implementing QRISs for more than a decade to promote, support, and incentivize quality improvement for ECE programs and to make ECE program quality more transparent to parents and funders. The federal Race to the Top—Early Learning Challenge (RTT-ELC) grants further expanded states' use of QRISs by giving funding priority to states implementing QRISs and by requiring that states undertake an independent evaluation of their QRISs.

Delaware's QRIS, known as Delaware Stars for Early Success (Delaware Stars for short), was first piloted in 2007 and gradually expanded thereafter. (See Text Box S.1 for an overview of Delaware Stars.) In 2011, Delaware was in the first group of states to receive an RTT-ELC grant. Like the QRISs in other states, Delaware Stars is motivated by a logic model which posits that children in higher-quality ECE programs will experience greater learning and development, particularly low-income children or those facing other disadvantages. Thus, an underlying premise of the QRIS is that the summary rating of quality—which begins in Delaware Stars with Starting with Stars (the first rating level) and continues from Star 2 to Star 5—measures meaningful differences in quality as ECE programs rise from the lowest quality rating to each successively higher rating.

Text Box S.1. Key Features of Delaware Stars

- **Voluntary:** Delaware Stars is a voluntary QRIS open to licensed ECE programs in good standing, including small and large family child care (FCC) providers and centers, as well as license-exempt school-based preschool programs.
- **Five rating tiers:** Programs enter Delaware Stars at the Starting with Stars level and move to Star 2 after completing an orientation session, a six-hour "Building on Quality" session, a visit from the state's TA provider, and a completed initial Quality Improvement Plan (QIP). Star 3 through Star 5 statuses are achieved through a verification process that involves observation of program quality using the Environment Rating Scale (ERS) and verification that a program meets a set of essential quality standards and other nonessential standards in four domains. Each rating tier from Star 3 to Star 5 requires that programs meet a higher ERS threshold, additional essential standards, and a higher point total for all other standards.
- **Alternative pathway programs:** Programs accredited through the National Association for the Education of Young Children (NAEYC) automatically receive a Star 5 rating. Stand-alone Head Start, Early Childhood Assistance Program (ECAP), and Part B, Section 619, public school preschool programs enter at Star 4 and advance to Star 5 by meeting the ERS threshold only.
- **Financial incentives:** Depending on the rating tier, programs are eligible for financial incentives to help improve or maintain quality. These include Quality Improvement (QI) Grants (Star 2 to Star 4 programs); Infrastructure Fund grants (Star 2 to Star 5 programs); Purchase of Care (POC) Tiered Reimbursement Bonuses (Star 3 to Star 5 programs); Compensation, Retention, and Education (CORE) Awards (for staff in Star 3 to Star 5 programs); and Teacher Education and Compensation Helps (T.E.A.C.H.) Early Childhood® scholarships.
- **Quality improvement supports:** Programs in Delaware Stars receive technical assistance (TA) support based on their rating level and QIP. Both general and specialized TA is available. In addition, Stars Plus is a cohort-based program open to Delaware Stars providers with a large proportion of children receiving POC subsidies that offers even more intensive TA and other supports for quality improvement.

Given that the rating structures of most QRISs are a complex aggregation of an array of quality standards or indicators, it has been important to understand whether the rating system in any given QRIS captures quality differences as intended. Such analyses are designed to validate the ratings component of the QRIS. Validation studies may examine the validity of the underlying concepts in the rating scale, assess the measures used in the rating scale and their psychometric properties, examine the relationship between the rating results and other validated quality measures, or establish whether higher rates are associated with better child developmental outcomes. The ultimate goal of such validation studies is to determine whether the rating scale identifies meaningful differences in program quality. As such, a validation study is one component of a larger evaluation of a QRIS and its performance.

In the case of Delaware Stars, a large group of stakeholders designed and piloted the QRIS, selecting standards that research had indicated were important for the health, safety, and development of children. To obtain a Star 3 to Star 5 rating (the top three rating tiers), Delaware ECE programs must achieve a given threshold on the relevant environment rating scale (ERS) for their program type. They also accumulate points from among more than 30 standards that are clustered within four quality domains—Family and Community Partnerships, Qualifications and Professional Development, Management and Administration, and Learning Environment and Curriculum—with a specified minimum required number of points to obtain a given Stars rating. While ECE providers do have some choice in which standards they seek to meet, Delaware has modified this system several times, making the system successively more demanding. The latest change, which started to take effect during this evaluation, was to make up to six standards mandatory to reach Star 4 or Star 5.

In light of the complexity and breadth of the Delaware QRIS, two key questions are as follows:

- Do ECE programs with higher ratings in the QRIS deliver higher-quality care and early learning than those with lower ratings?
- Do children in programs with higher ratings in the QRIS have better learning and developmental outcomes than children in programs with lower ratings?

As shown in Table S.1, these two validation questions are central to RAND's evaluation of Delaware Stars, which has spanned three years and three reports. Specifically, the evaluation addresses questions about the validity of the QRIS ratings (Q1 to Q4 in Table S.1), as well as questions about the performance of other key components of the QRIS, such as the efficacy of the Delaware Stars improvements supports (which are the "I" in the QRIS), the adequacy of the financial incentives, and the performance of other aspects of the system (Q5 to Q9 in Table S.1). In this third and final report, our primary focus is on the first four questions in Table S.1 relating to the validation of the Delaware Stars ratings, although we also bring new information to bear on the remaining evaluation questions. Taken together, the three reports address the full list of questions in Table S.1.

In the remainder of this summary, we first provide a brief overview of our approach to the evaluation. We then turn to our findings in regard to the questions in Table S.1. These findings, summarized in Table S.1, are based on the latest analyses detailed in this report, as well as results from the Year 1 report and Year 2 report.

Table S.1. Study Research Questions and Key Findings

Study Research Questions	What We Found
Delaware Stars Ratings and Program Quality	
Q1. To what extent do the quality tiers of Delaware Stars accurately reflect differential levels of program quality such that programs at the top levels provide a higher-quality care and early learning experience than programs at lower levels?	• Alternative quality measures rose modestly with Delaware Stars ratings, but the increases were usually statistically insignificant and small in magnitude. • For the most part, alternative quality measures were not positively related to scores on the five components that make up the overall Delaware Stars rating (i.e., ERS and the four quality domains in which points are obtained). • There were no statistically significant relationships between the essential standards and alternative quality measures.
Q2. What is the relationship between program characteristics and quality in Delaware Stars?	• FCCs had high quality scores on some measures, but this may reflect sample selectivity. • Programs serving majority ECAP children had lower instructional support quality, on average.
Delaware Stars Ratings and Child Developmental Outcomes	
Q3. All else being equal, do young children participating in higher-rated programs have better learning and developmental outcomes than similar children in nonparticipating or lower-rated programs?	• Differences in children's development across rating tiers were generally small and statistically insignificant. • Children in Star 5 center-based programs modestly outperformed children in Starting with Stars and Star 2 programs on executive function skills, an important predictor of academic success. • There was no evidence that top-rated Delaware Stars center-based programs conferred greater benefits for children from low-income families as compared with lower-rated programs.
Q4. What dimensions of Delaware Stars program ratings are most vital to child learning and developmental outcomes?	• For center-based programs, points obtained in two quality domains—Management and Administration and Qualifications and Professional Development—were associated with higher scores on some assessments of early academic and cognitive skills. • Children in center-based programs that met more of the six essential standards modestly outperformed children in programs that met fewer essential standards in terms of executive function skills.
Delaware Stars System and Program Quality Improvement	
Q5. To what extent does the TA (i.e., on-site support, orientation, etc.) provided to Delaware Stars participants help providers to move up in Delaware Stars?	• TA was the second-largest financial investment made by Delaware Stars (after POC Tiered Reimbursement Bonuses). • Delaware Stars programs received the intended number of TA visits. • Directors highly valued TA overall, especially the specialty TA and specialized training designed to improve program quality in targeted areas. • Directors identified the need for greater consistency in TA guidance, more frequent TA visits, and less frequent reassignments of TA staff.
Q6. To what extent do high-need programs that participate in Stars Plus enhance their program quality?	• Directors in Stars Plus viewed the peer network and the extra TA they received as valuable supports for improving quality.

Table S.1. Study Research Questions and Key Findings, Continued

Study Research Questions	What We Found
Delaware Stars System Performance	
Q7. Are the financial incentives and supports for providers sufficient to support the needed quality improvements?	• The five types of financial incentives represent a large investment overall and conferred an average of over $27,000 per program in 2013–2014. • Directors value these incentives as supports for making and sustaining program quality improvements, especially the POC Tiered Reimbursement Bonuses, which were the largest incentive measured by total dollars as of 2013–2014.
Q8. What do consumers understand about Delaware Stars? Do consumers ask about ratings?	• Directors reported using Delaware Stars marketing materials but also noted low parental awareness of Delaware Stars and saw low marketing value for the program.
Q9. How well do the Delaware Stars system components operate?	• Administrative data show continued recruitment of programs into Delaware Stars and upward movement among those in the system, with especially high participation rates among centers. • Time and resource constraints were the most-cited barriers to moving up the Delaware Stars rating tiers. • Directors viewed several of the essential standards as the most difficult to attain. • Improving quality is a top motivator for participating in Delaware Stars. • Directors identified multiple benefits for their staff from being in Delaware Stars.

Evaluation Approach

The RAND evaluation has proceeded in two phases, with a set of initial analyses that could be accomplished with existing data or qualitative information examined in the first two evaluation reports, followed by more intensive primary data collection, the focus of this final report.

As noted earlier, Delaware Stars has undergone a series of modifications since it was first piloted in 2007. In fact, it was continuing to change during the course of the three-year evaluation. As a consequence, a majority of the programs we studied were last rated under the prior Delaware Stars structure, before OEL had mandated that programs meet certain standards to reach the top two rating tiers. Thus, this evaluation primarily reflects a maturing Delaware Stars QRIS before its fullest implementation. It sets a baseline against which to compare future studies of Delaware Stars when participating programs have all been rated against the latest, more stringent set of standards.

Initial Investigations

To set the stage for the design and implementation of primary data collection on providers and children in Delaware in support of the validation study, our Year 1 report included a literature review to understand prior QRIS validation research; an analysis of administrative data to document provider participation in Delaware Stars and the quality components that make up the QRIS ratings; interviews and focus groups with system administrators, providers, and families to

understand their experience with Delaware Stars; and an analysis of existing national data to explore the relationship between the dimensions of quality in the Delaware Stars QRIS and child developmental outcomes. In the Year 2 report, we provided updated analyses of provider participation using Delaware Stars administrative data and conducted new analyses of Delaware Stars financial incentives and TA, also based on administrative data. These analyses primarily served to address Q5 through Q9 in Table S.1.

Primary Data on Providers and Children for Validation Study

The main focus of this final report is on the first four research questions in Table S.1, which make up the validation component of the evaluation. To answer these questions, we collected data for a representative sample of 181 licensed ECE programs in Delaware (out of approximately 1,200 such providers)—small and large FCC providers, centers, and school-based preschools—and the children they enrolled during 2014–2015. Most providers were in Delaware Stars when first sampled or they entered Delaware Stars during the field period, although some were not enrolled during the period of data collection. The data consist of the following components:

- *interviews of program directors in the provider sample* in the fall of 2014 and spring of 2015, referred to as the *director interviews*, to collect information about their knowledge of and experience with the Delaware Stars system
- *observational assessments of program quality* in the 2014–2015 program year for the provider sample through observations of up to three classrooms for each provider (or groups in the case of FCCs) using multiple measures of quality that have been validated in other research, to capture dimensions of quality rated in Delaware Stars (especially for the program quality assessment [PQA]), but are otherwise not used to determine Delaware Stars ratings (see Table S.2)
- *developmental assessments for a child sample* of approximately 1,100 toddlers and preschool-age children enrolled in the observed large FCCs and center- and school-based classrooms in the provider sample, with assessments in the fall of 2014 and the spring of 2015 of multiple measures of early academic and cognitive skills, as well as social-emotional and behavioral development (see Table S.2).

Child and family background characteristics were also collected using a brief self-administered parent survey and through information collected from the child's ECE provider.

The data collected for providers on observed quality and the child developmental assessments provide the basis for analyses of Q1–Q4 in Table S.1. The director survey provided relevant information for most of the other study questions to supplement what had been learned from the analyses presented in the first two reports.

Table S.2. Measures of ECE Program Quality and Child Development Collected for the Provider and Child Samples

Measurement Tool	What It Measures
Provider Sample Program Quality	
Preschool Program Quality Assessment (PQA)—Second Edition	Global measure of quality collected by classroom observation and interview (teacher, director)
Classroom Assessment Scoring System (CLASS)	Measure of teacher-child interactions collected by observation for a classroom; for use in center- or home-based setting
Arnett Caregiver Interaction Scale (CIS)	Measure of teacher-child interactions collected by observation at the teacher level; for use in center- or home-based settings
Child Sample Developmental Assessments	
Peabody Picture Vocabulary Test (PPVT)	Direct assessment of ability to understand spoken words
Woodcock-Johnson III (WJ) Letter Word Identification	Direct assessment of early reading skills
Woodcock-Johnson III (WJ) Applied Problems	Direct assessment of skill in solving practical math problems
Head-Toes-Knees-Shoulders (HTKS)	Performance-based assessment of executive function (e.g., attention, inhibitory control, and working memory)
Devereaux Early Childhood Assessment (DECA)	Teacher-rated assessment of • protective factors (overall positive social and behavior skills) • behavioral concerns (incidence of emotional and behavioral problems)

What We Found

To answer the first four questions listed in Table S.1, which is the validation portion of the overall evaluation, we followed the most recent generation of QRIS validation studies in using other validated measures of program quality to compare with Delaware Stars ratings and examined the relationship between program quality ratings and child developmental outcomes. For the latter estimation, we accounted for children's initial level of skill (as of fall 2014) and controlled for other child and family background characteristics that might also explain children's learning, such as parent education levels and family income. The analyses of child outcomes primarily generalize to center-based programs where 98 percent of the children in the sample were enrolled. Analyses from the first two study reports and the director interview collected in 2014–2015 provide the basis for addressing the other questions in Table S.1.

Delaware Stars Ratings and Program Quality

- **Alternative measures of program quality increased modestly as Delaware Stars rating levels rose.** Average scores for the alternative quality measures collected for the provider sample generally increased as Delaware Stars rating levels increased from Star 3 to Star 5, although at a modest rate and generally not with statistically significant increases from one star level to the next. The exceptions were statistically significant but small increases from Star 3 to Star 5 in average PQA scores—the most comprehensive alternative quality measure that captured the key quality constructs rated in Delaware

Stars—and in average CLASS Pre-K Instructional Support scores—a key measure of teacher-child interactions that is implicitly measured in the ERS used to generate the Delaware Stars rating. At the same time, the level of quality as measured by the PQA and CLASS, even for Star 5 programs, was below the level expected for high-quality programs, and the incremental improvements in quality in moving from Star 3 to Star 5 were small according to these measures.

- **For the most part, alternative measures of quality were not positively related to scores on the components that make up the Delaware Stars ratings.** We generally did not observe the expected stair-step increase in the alternative quality measures as the number of points a program obtained in a given quality domain of Delaware Stars rose or the ERS score increased. The expected pattern was evident and statistically significant between the CLASS Pre-K Instructional Support subscale and the Qualifications and Professional Development domain and between the PQA and the Learning Environment and Curriculum domain. Further, we found no statistically significant relationships between the essential standards and our alternative measures of program quality.
- **Some program characteristics were related to the alternative measures of quality.** Small FCCs scored higher on average than centers on the PQA and the Emotional Support CLASS scale, but this may reflect the selectivity of the FCC providers in our sample. Programs in which a majority of enrollees received ECAP assistance scored lower on the CLASS Instructional Support subscale.

Delaware Stars Ratings and Child Developmental Outcomes

- **Differences in children's development across rating tiers were generally small and statistically insignificant.** For the children in our sample, we found some developmental measures—both early academic skills and social-emotional and behavior skills—where average performance was higher in Star 4 and Star 5 programs compared with children in Star 3 programs. However, we had a relatively small sample of children in Star 3 programs, so we cannot conclude that the differences we observed across rating levels were true differences or were specific to the sample of children for whom we collected data.
- **Children in Star 5 center-based programs modestly outperformed children in Starting with Stars and Star 2 programs on executive function skills.** The difference in performance by Delaware Stars rating levels was only statistically significant in the case of the HTKS, a measure of executive function which is itself an important predictor of academic success, and then only between Starting with Stars and Star 2 programs and Star 5 programs (a difference of 0.34 standard deviations).
- **Children from low-income families did not experience differential learning in higher-rated programs when compared with lower-rated programs.** We defined low-

income children as those children whose families had incomes of $25,000 or less or those receiving subsidized care through Head Start, ECAP, or POC. We generally found no difference in the scores of children from low-income families in low- versus high-rated programs.

- **Two quality domains of Delaware Stars were predictive of modest differences in selected children's outcomes.** Scoring high on the Qualifications and Professional Development domain in Delaware Stars was related to modestly higher average WJ–Letter Word Identification scores and WJ–Applied Problems scores, with differences in performance between lower-scoring and higher-scoring programs equal to 0.18 and 0.25 standard deviations, respectively. We also found a nonlinear but statistically significant relationship between Management and Administration scores and children's performance on the HTKS measure of executive function, with a contrast in performance of 0.32 standard deviations between programs scoring in the lowest quartile and the highest quartile of the domain.
- **Children in programs that met more of the six essential standards modestly outperformed children in programs that met fewer essential standards in terms of executive function skills.** The sum of the points programs obtained on each of the six essential standards was predictive of moderate differences of approximately 0.33 standard deviations on the measure of executive function. Separately, none of the six essential standards was associated with children's outcomes after accounting for multiple hypothesis testing.

Delaware Stars System Performance

- **Delaware Stars makes a substantial investment in TA, and TA is highly valued, although there may be room for improvement.** Administrative data documented high participation rates in TA, at a frequency consistent with the planned level of support. In survey responses, directors were strongly enthusiastic about the TA supports within Delaware Stars, especially specialty TA and specialized training designed to improve quality in targeted areas. Those in Stars Plus also endorsed the supports of having a peer network of other directors with whom to share ideas and lessons. Directors saw room for improving TA through greater consistency in TA guidance, more frequent visits, and lower turnover among TA staff assigned to their program.
- **Financial incentives represent another area of significant investment for Delaware Stars, and directors view the funds as valuable for improving quality.** Delaware Stars system data document the considerable investment made in financial incentives, both at the system level and in terms of the combined value of the incentives on a per-program basis. In survey responses, directors indicated that they highly value the financial supports offered through Delaware Stars, most particularly the POC Tiered

Reimbursement Bonuses, which constitute the largest share of the Delaware Stars financial incentives.

- **Low parental awareness of Delaware Stars limits value to providers from participation.** Directors corroborated what our focus groups of parents (in the Year 1 report) found: Parents have low levels of knowledge about Delaware Stars. Most directors reported that they were supplying information about their participation in Delaware Stars and their rating, but few reported that parents inquired about ratings or appeared to make decisions about provider choice with ratings in mind. Even though a majority of providers wanted to improve quality to attract more families, most providers did not think that they were benefiting from Delaware Stars in terms of their marketability and enrollments. In other words, ratings may not be serving as a strong market signal.
- **Programs face challenges in advancing through the rating tiers, but they are motivated to improve quality.** While most directors in Delaware Stars support the goal of moving to higher quality, they also indicated that the most challenging standards to meet are those that Delaware Stars has newly mandated to reach Star 4 and Star 5. Standards related to staff credentials, curriculum, and child assessments were viewed as among the hardest to meet. Beyond the motivation to improve quality, directors saw multiple areas of benefit for their staff from participation in Delaware Stars, such as professionalization of teaching. At the same time, for a small minority of programs, participation in Delaware Stars was described as a stressful experience for program staff.

Findings in the Context of Other QRIS Validation Studies

Our results for Delaware Stars are entirely consistent with what has been found for other QRIS validation studies. Although most validation studies that examine ratings relative to other measures of quality find the expected positive relationships, the correlations are generally weak. The increase in the average level of provider quality in moving from rating tier to rating tier is small by comparison to the implied movement embedded in the rating scale (such as a one-scale-point increase in the ERS required to move up a rating tier in Delaware Stars). Likewise, the lack of a strong relationship between children's development and QRIS ratings is a common finding. Finally, in many cases, the absolute level of program quality, even at the highest rating tier, has not been at the level of programs with demonstrated impact on children's development.

It is also important to recognize that QRISs have gained currency as a mechanism for ECE quality verification and program improvement at a time when evidence is accumulating that the available measures of quality may not be as strongly related to children's outcomes as suggested by earlier research. Various nonexperimental and quasiexperimental research has found either a small difference or no difference in children's outcomes based on such structural measures of quality as materials in the classroom or teacher-to-student ratios. Even such process measures as

the quality of teacher-student interactions (for example, the CLASS measures) do not consistently predict children's development, although subscales of measures may be more predictive.

Finally, we note that QRIS validation studies, like our study of Delaware Stars, do not provide an evaluation of the impact of the QRIS on child outcomes. Our empirical findings indicate that we are not detecting a strong positive relationship between Delaware Stars quality ratings and the alternative measures of quality that we gathered or children's learning and developmental outcomes. That is not the same thing as saying that there have been no improvements in child outcomes as a result of implementing Delaware Stars. Even if the Delaware Stars rating structure does not differentiate program quality as strongly as would be desired, if (1) ECE program quality is improving over time in ways that favorably affect children's development and (2) that improvement is happening for much of the ECE provider base, we would expect that at least some children will have experienced better developmental outcomes as a result of the implementation of Delaware Stars compared with a status quo in which the QRIS had not been implemented.

Limitations

Although we implemented rigorous methods to ensure the quality of the data collected and adopted state-of-the-art methods employed in the most recent generation of QRIS validation studies, there are a number of limitations that are relevant when interpreting our findings with respect to the first four study questions. In brief, these include the following:

- As mentioned previously, we have not assessed the validity of the fully phased-in rating structure of Delaware Stars. This is because during the period covered by our data collection, most programs in Delaware Stars with a Star 3 to Star 5 rating had yet to have their ratings determined using the newest ratings structure, particularly the use of essential standards, which were not fully phased in until July 2016.
- We had considerably lower response rates for FCC providers compared with center- or school-based programs. Thus, our findings are most relevant for school- and center-based providers—the ECE settings where most children in Delaware, especially preschoolers, are enrolled. By design, our study did not examine care quality or developmental outcomes for infants in center- or home-based care settings.
- As with other validation studies, our analyses of the relationship between program quality and children's developmental outcomes are potentially affected by the role that parental choice plays in which programs children attend. Although our empirical models accounted for a number of relevant child and family background characteristics, we cannot rule out possible selectivity bias. At the same time, if we expect children from more advantaged families to attend higher-quality programs, the selection bias would

make us more likely to find a positive relationship between program quality and child development. Given that we did not find such a relationship, the issue of selectivity bias does not appear to be a concern.

- As with other similar validation studies, our analyses are affected by the limitations of the measures of program quality and child development that we employed. This issue is a general challenge for the field, as the premise for a QRIS is that we have measures of quality that can be used to detect meaningful differences in program quality that can then be incorporated into the rating structure.

Despite these concerns, the evaluation of Delaware Stars is an important step in the process of implementing and sustaining an effective QRIS. This evaluation represents a baseline assessment of Delaware Stars that can serve as a reference point as Delaware Stars evolves further and subsequent evaluations are conducted.

Recommendations for OEL and Other Stakeholders

The evaluation of Delaware Stars has provided independent, objective, and rigorous empirical evidence of the extent to which rating tiers reflect relevant differences in the quality of home- and center-based providers and documented providers' experience with TA, financial supports, and other QRIS features. Based on our findings and those of other QRIS validation studies and with the study limitations in mind, we offer several recommendations for the Delaware Office of Early Learning (OEL) and other stakeholders to implement and sustain an effective ECE QRIS.

- **Learn from other QRIS validation studies.** With nearly two dozen states involved in QRIS evaluations, OEL should look beyond the findings of any one study for any particular QRIS and discern the broader findings with relevance for QRIS design and implementation. Research syntheses are needed—and perhaps formal meta-analyses as well—to determine why some state rating structures appear to capture quality differences while others do not. This body of research evidence may also shed light on other aspects of system performance, such as the role that the nature and intensity of quality improvement supports and financial incentives play in advancing classroom practices and program quality.
- **Consider further refinements to the Delaware Stars rating structure.** We recommend that OEL consider a simplified, streamlined rating system based on carefully selected measures of the dimensions of quality that appear to matter most for achieving the goals of the QRIS, such as improvements in child developmental outcomes. If a global quality scale (like the ERS or PQA) is to be used at all, then OEL should consider using subscales of existing measures like the CLASS instructional support that research has shown to be most strongly predictive of children's development. OEL should also consider reducing the more than 30 standards within Delaware Stars to a smaller number

that are either essential for children's outcomes or are deemed essential for any other goals of the QRIS (e.g., financial stability). Another direction would be to raise the quality standards required to reach the highest rating level, so that programs that reach Star 5 consistently perform at the highest quality levels. While further modification to Delaware Stars carries a cost, a simplified system should result in savings that can be redirected elsewhere in the system and should produce a welcome reduction in the burden of the QRIS on providers and their staff.

- **Strengthen quality improvement supports in Delaware Stars.** As Delaware Stars has continued to incorporate new center- and home-based providers, there is an ongoing need to support providers in their efforts to improve quality and advance toward higher ratings. Although providers are generally positive about current TA supports and financial incentives, there is room for improvement. With respect to TA in particular, there is scope for lowering TA caseloads, increasing TA training, and improving the guidance that TAs provide. Future research could assess the effectiveness of TA and other quality improvement supports, as well as the impact of various types of financial incentives.
- **Strengthen the marketing of Delaware Stars to families.** Our findings suggest that there is opportunity for further marketing of Delaware Stars to families with young children, an important centralized role for OEL. OEL should take stock of current marketing activities and evidence of their impact, compare those strategies with best-practice guidance in the field (including successful strategies in other states), and determine where new approaches may be called for. Any marketing activities and their timing would need to account for any further planned modifications to Delaware Stars.
- **Enhance administrative data systems to support ongoing system monitoring and quality improvement.** We recommend that OEL invest in a data manager to revise the Delaware Stars database both by reducing the number of data elements collected and by improving the accuracy of the data elements that remain. Ideally, a linked database of licensed programs to Delaware Stars data and Head Start and ECAP enrollments will include refreshed enrollments on a quarterly basis and will integrate financial incentives and TA data to allow OEL to monitor the outputs of the QRIS and evaluate its performance. Linked child-level data can also be used to evaluate the effects of participation in higher-rated programs versus lower-rated ones on kindergarten readiness and subsequent school performance. Such analytics are made possible by robust integrated data systems, and OEL should continue investments in this area prompted by earlier findings from this evaluation. This means ensuring that essential data elements are captured in a standardized, timely way.

Taken together, these recommendations provide a series of action steps that comport with the growing recognition of the importance of using data, analytics, evidence, and evaluation—

known as the “moneyball” approach—to provide a regular feedback mechanism for assessing the current landscape, identifying where improvements are needed, implementing the needed modifications, and then monitoring and evaluating further. In this way, OEL can model a culture of learning and improvement that ideally would permeate all levels of the early learning system, from the micro level (e.g., teachers, classrooms, and providers) to the broader ECE system level (e.g., child care licensing, QRIS, professional development system).

Acknowledgments

The Delaware OEL enthusiastically supported this evaluation, which started under the leadership of Harriet Dichter, continued under Evelyn Keating, and concluded under the direction of Susan Perry-Manning. Ariel Ford served as our main point of contact in the final phase of the study. We remain grateful for the continued help of Kristy Sheffler and Annette Searfoss at the University of Delaware Institute for Excellence in Early Childhood to fact-check program changes to Delaware Stars over time. Other staff from OEL who provided input included Christine Anderson, Caitlin Gleason, and Carmen Gallagher.

Over the course of the three-year evaluation, we benefited from Delaware-based members of the project advisory group. These local advisory group members included Elaine Archangelo, Delaware Department of Health and Social Services; Madeleine Bayard, Rodel Foundation of Delaware; Melissa Browne, Latin American Community Center; Rena Hallam, University of Delaware; Vicky Kelly, Delaware Department of Services for Children, Youth and Their Families; Andrea Prettyman, Children and Families First; Helen Riley, St. Michael's School; Michelle A. Taylor, United Way of Delaware; and Kathy Wilson, Delaware Department of Education. Informal feedback was also received from David Blowman, Department of Education; John Fisher-Klein, Newark Day Nursery; Deborah Gottschalk and Belvie Herbert, Department of Health and Social Services; Paul Herdman, Rodel Foundation of Delaware; Kim Lowman, Kids Count; Leslie Newman, Children and Families First; Michelle Shaivitz, Delaware Association for the Education of Young Children; and Betty Gail Timm, Office of Child Care Licensing.

Our evaluation benefited as well from the input of the following early childhood research experts also on our advisory committee: Miriam Calderon, BUILD Initiative; James Elicker, Purdue University; Kelly L. Maxwell, Child Trends; Deborah (Debi) Mathias, BUILD Initiative; Geoffrey A. Nagle, Erikson Institute; and Kathy Thornburg, AEM Corporation.

We also extend our gratitude to the center- and home-based program directors who participated in our study and to the parents of children and the children themselves who consented to participate.

Finally, we benefited from thorough technical reviews provided by Jill Cannon of RAND and by Katherine Magnuson at the University of Wisconsin–Madison. At RAND, we received valuable programming support from Teague Ruder and administrative support from Sarah Cressy. We also acknowledge Nora Spiering, who edited the summary and report, and Julienne Amsden, who managed the document production.

Abbreviations

ACS	American Community Survey
CDA	Child Development Associate
CIS	Caregiver Interaction Scale
CLASS	Classroom Assessment Scoring System
CORE	Compensation, Retention, and Education
DECA	Devereaux Early Childhood Assessment
DOE	Department of Education
ECAP	Early Childhood Assistance Program
ECE	early care and education
ECERS-R	Early Childhood Environment Rating Scale–Revised
ERS	environment rating scale
FCC	family child care
FCCERS-R	Family Child Care Environment Rating Scale–Revised
HTKS	Head-Toes-Knees-Shoulders
ITERS-R	Infant/Toddler Environment Rating Scale–Revised
NAEYC	National Association for the Education of Young Children
OCCL	Delaware Office of Child Care Licensing
OEL	Delaware Office of Early Learning
PLBS	Preschool Learning and Behavior Scale
POC	Purchase of Care
PPVT	Peabody Picture Vocabulary Test
PQA	Program Quality Assessment
preLAS	Preschool Language Assessment Scale
PUMS	Public Use Microdata Sample
QI	quality improvement
QIP	quality improvement plan

QRIS	quality rating and improvement system
RTT-ELC	Race to the Top—Early Learning Challenge
SCBE-30	Social Competence and Behavior Evaluation short form
TA	technical assistance
T.E.A.C.H.	Teacher Education and Compensation Helps
TERA–3	Test of Early Reading Ability, Third Edition
TOPEL	Test of Preschool Early Literacy
TVIP	Test de Vocabulario en Imagenes Peabody
WJ	Woodcock Johnson

Chapter 1. Introduction

Delaware's quality rating and improvement system (QRIS), called Delaware Stars for Early Success (hereafter Delaware Stars), began in 2007. The main purpose of the QRIS is to raise the quality of early care and education (ECE) and school-age programs within the state and improve child developmental outcomes (State of Delaware, 2011).[1] With funding from a federal Race to the Top—Early Learning Challenge (RTT-ELC) grant, the Delaware Office of Early Learning (OEL) set out in December 2011 to further improve the quality of ECE programs and to increase enrollment of children from birth to age five in programs that are highly rated within the Delaware Stars system.[2] A requirement of an RTT-ELC grant is an evaluation of the QRIS, and the Delaware OEL contracted with RAND to conduct that evaluation.

RAND's evaluation addresses research questions in four main areas:

- *Delaware Stars ratings and program quality:* Do ECE programs with higher ratings in the QRIS deliver higher-quality care and early learning than those with lower ratings? What is the relationship between program characteristics and quality in Delaware Stars?
- *Delaware Stars ratings and child developmental outcomes:* Do children in programs with higher ratings in the QRIS have better learning and developmental outcomes than children in programs with lower ratings? What dimensions of Delaware Stars program ratings are most vital to child learning and developmental outcomes?
- *The Delaware Stars system and program quality improvement:* Does technical assistance (TA) help providers meet the standards and move up in Delaware Stars?
- *Delaware Stars system performance:* Are the financial incentives and supports for providers sufficient to support the needed quality improvements? What do parents, as consumers, understand about Delaware Stars? How well do the Delaware Stars system components operate?

We refer to the first two research issues as the primary components of a validation study of Delaware Stars, in keeping with other studies of QRISs that seek to understand whether QRIS ratings capture meaningful differences in program quality that are important for children's

[1] Throughout the report, we use the terms *program* and *provider* interchangeably when referring to a center or FCC setting.

[2] To date, there have been a total of three rounds of RTT-ELC grants, ranging in size from $30 million to $75 million per state, awarded to a total of 20 states, including Delaware. In alphabetical order, these 20 states and the phase in which they were awarded an RTT-ELC grant are as follows: California (phase 1), Colorado (phase 2), Delaware (phase 1), Georgia (phase 3), Illinois (phase 2), Kentucky (phase 3), Maryland (phase 1), Massachusetts (phase 1), Michigan (phase 3), Minnesota (phase 1), New Jersey (phase 3), New Mexico (phase 2), North Carolina (phase 1), Ohio (phase 1), Oregon (phase 2), Pennsylvania (phase 3), Rhode Island (phase 1), Vermont (phase 3), Washington (phase 1), and Wisconsin (phase 2).

development. The third and fourth topics constitute additional research questions that encompass the broader evaluation of Delaware Stars.

In the remainder of this introduction, we first provide additional detail on the Delaware Stars evaluation components and approach. We then provide an overview of the Delaware Stars system. To provide context for the evaluation of Delaware Stars, we compare Delaware with other states in terms of child care licensing and QRISs. We also highlight findings from evaluations of other QRISs, particularly other validation studies that have aimed to address the first two research topics listed above. A final section provides a road map for the remainder of the report.

Research Questions and Approach

Table 1.1 lists the detailed research questions of interest in the Delaware Stars evaluation. For each question, we indicate whether findings are available in one of the two earlier reports from the evaluation (see Text Box 1.1) or in this final report.

Table 1.1. Study Research Questions and Reporting of Findings

Study Research Questions		Where Findings Are Reported	
		Earlier Reports	This Report
Delaware Stars Ratings and Program Quality			
Q1.	To what extent do the quality tiers of Delaware Stars accurately reflect differential levels of program quality such that programs at the top levels provide a higher-quality care and early learning experience than programs at lower levels?		Chapter 3
Q2.	What is the relationship between program characteristics and quality in Delaware Stars?		Chapter 3
Delaware Stars Ratings and Child Developmental Outcomes			
Q3.	All else being equal, do young children participating in higher-rated programs have better learning and developmental outcomes than similar children in nonparticipating or lower-rated programs?		Chapter 4
Q4.	What dimensions of Delaware Stars program ratings are most vital to child learning and developmental outcomes?		Chapter 4
Delaware Stars System and Program Quality Improvement			
Q5.	To what extent does the TA (i.e., on-site support, orientation, etc.) provided to Delaware Stars participants help providers to move up in Delaware Stars?	Year 2 Report	Chapter 5
Q6.	To what extent do high-need programs that participate in Stars Plus enhance their program quality?	Year 1 and Year 2 Reports	Chapter 5
Delaware Stars System Performance			
Q7.	Are the financial incentives and supports for providers sufficient to support the needed quality improvements?	Year 2 Report	Chapter 5
Q8.	What do consumers understand about Delaware Stars? Do consumers ask about ratings?	Year 1 Report	Chapter 5
Q9.	How well do the Delaware Stars system components operate?	Year 2 Report	Chapter 5

Text Box 1.1. Focus of First Two Evaluation Reports

For the Year 1 report from the project (Schwartz et al., 2014), we conducted a literature review, held focus groups of parents and administrators, and analyzed state and national data to address the following topics and associated questions:

- *Placing the evaluation in context*: What do we know from prior QRIS validation research? What lessons can inform the design of the Delaware Stars evaluation?
- *Participation in Delaware Stars:* What is the extent of participation in Delaware Stars on the part of home- and center-based providers, and what do the ratings reveal about dimensions of program quality?
- *Experience with Delaware Stars:* What are the experiences of system administrators, providers, and families with Delaware Stars? What aspects of the system are working well, and what challenges do stakeholders identify?
- *Quality features and child outcomes:* What can we learn from existing national data about the relationship between the dimensions of quality in the Delaware Stars QRIS and child developmental outcomes?

In the Year 2 report (Auger, Karoly, and Schwartz, 2015), we provided an updated set of analyses of Delaware Stars participation and quality rating outcomes included in the first report, again based on administrative data. We expanded upon the earlier analyses by examining administrative data on the financial incentives of Delaware Stars and the TA provided to participating programs. Specifically, in the Year 2 report we addressed the following topics and specific questions:

- *Program participation and quality ratings:* Are more programs and children participating in Delaware Stars in October 2014 as compared with January 2014? Is the distribution of providers across counties and star levels similar at both time points? Are programs advancing at similar rates in the rating system?
- *Financial incentives:* Are providers using the available financial incentives? What financial incentives do providers participate in, and what is the value of the financial incentives they receive? How do participation in financial incentives and the value of those incentives vary with Delaware Stars ratings and other provider characteristics?
- *TA:* How many visits from a TA provider are programs receiving on average? What is the typical duration of a visit? Does the length and number of visits differ by provider characteristics?

The primary focus of this third and final report is the first four questions (Q1 to Q4). To answer them, we report on analyses based on primary data collected for a sample of 181 licensed small and large family child care (FCC) providers, licensed centers, and school-based preschools—including providers participating and not participating in Delaware Stars—and their enrolled children. In particular, the primary data consist of the following components:

- *interviews of program directors in the provider sample* in the fall of 2014 and spring of 2015, referred to as the *director interviews*, to collect information about their knowledge of and experience with the Delaware Stars system
- *observational assessments of program quality* in the 2014–2015 program year for the provider sample through observations of up to three classrooms per provider (or groups, in the case of FCCs) using multiple measures of quality

- *developmental assessments for a child sample* of approximately 1,100 toddlers and preschool-age children enrolled in the observed large FCCs and center- and school-based classrooms in the provider sample, with assessments conducted in the fall of 2014 and the spring of 2015 on measures of early academic, cognitive, social-emotional, and behavioral development. Child and family background characteristics were also measured through a brief self-administered parent survey and information collected from the ECE program.

By drawing on findings from the two earlier reports, as well as data we collected for this final report, we are able to address other questions related the Delaware Stars system and program quality improvement (Q5 and Q6) and the performance of the Delaware Stars system (Q7 through Q9).

Description of Delaware Stars Rating and Quality Improvement System

In 2007, Delaware implemented a pilot QRIS statewide, and in 2009, the state legislature passed a bill authorizing the QRIS in state law (see Schwartz et al., 2014, for a complete overview). Participation by ECE providers in the QRIS is voluntary. To attract licensed ECE providers into the QRIS, incentives are available, such as higher levels of reimbursement for the child care assistance subsidy, known as Purchase of Care (POC) Tiered Reimbursement Bonuses, and compensation for ECE professionals. The rating system includes both ECE providers and school-age care providers, but because the focus of the RTT-ELC grant and the associated RAND evaluation is on ECE providers, we limit our discussion to those aspects of the QRIS pertaining to infants, toddlers, and preschool-age children. Additional detail on the system is provided in Appendix A and at the Delaware Stars website (Delaware Institute for Excellence in Early Childhood, undated).

A number of modifications have been made to Delaware Stars over time, including during the period covered by the RAND evaluation. We begin in the next section with a description of Delaware Stars as of September 2014, the period when our collection of provider and child data began. This is the system that our study examined. We then summarize several key changes to Delaware Stars that began to be phased in as of January 2015. We conclude this section by placing Delaware's QRIS approach within the context of the systems in other states.

Structure of Delaware Stars as of Fall 2014

All licensed ECE programs can take part in Delaware Stars, including small FCC providers (up to six enrolled children), large FCC providers (seven to 12 enrolled children), and private and public center-based ECE providers. Public schools, including charter schools, are license-exempt but are still eligible to take part in Delaware Stars. To be eligible to participate in Delaware Stars, programs that are not public schools must have been licensed for at least six months and be

in good standing with the Delaware Office of Child Care Licensing (OCCL), POC, and the Child and Adult Care Food Program.

Table 1.2 displays the requirements for early childhood providers to advance through the five star levels of the QRIS. With the exception of alternative pathway programs, all ECE programs that elect to enter Delaware Stars begin at the first star level, Starting with Stars. At the Starting with Stars level, a Delaware Stars TA meets with the program director to create a quality improvement plan (QIP) that outlines what quality standards the program will focus on to advance in the star-rating system. The beginning star level also requires providers to complete an orientation and professional development session. Once all those requirements are met, programs are eligible to move up to Star 2.

Table 1.2. Rating Tier Requirements for All Provider Types and Alternative Pathway Requirements

		Alternative Pathway Requirements		
Star Level	**Common Requirements for All Provider Types**	**Public School 619 (Part B) Preschool Programs**	**Stand-Alone Head Start/ ECAP**	**NAEYC-Accredited Programs**
Starting with Stars	• Eligible licensed provider	—	—	—
Star 2	• Orientation • "Building on Quality" session • TA visit • QIP	• Orientation	—	—
Star 3	• ERS ≥ 3.4 • 40 or more points summed across each of four domains	• ERS ≥ 3.4 • 40 or more points summed across each of four domains (based on portfolio review)	—	—
Star 4	• ERS ≥ 4.4 • 60 or more points summed across each of four domains	• ERS ≥ 4.4 • 60 or more points summed across each of four domains (based on portfolio review)	• Orientation	—
Star 5	• ERS ≥ 5.4 • 80–100 points summed across each of four domains	• ERS ≥ 5.4 • 80–100 points summed across each of four domains (based on portfolio review)	• ERS ≥ 5.4	• Orientation

SOURCE: Delaware Institute for Excellence in Early Childhood, undated.
NOTES: ERS = environment rating scale. — = not applicable. The ERS thresholds for each star level took effect on July 1, 2014. Also, as of that date, the ERS is calculated without the Personal Care Routines subscale. The Parents and Staff section on the Infant/Toddler Environment Rating Scale–Revised (ITERS-R) and Early Childhood Environment Rating Scale–Revised (ECERS-R), as well as the Parents and Provider section on the Family Child Care Environment Rating Scale–Revised (FCCERS-R), are also omitted from the ERS scoring. Public School 619 (Part B) preschool programs are publicly funded preschool programs for children with disabilities, ages three to five. The Early Childhood Assistance Program (ECAP) is the state-funded comprehensive child development program for four-year-olds from low-income families.

To advance from Star 2 to Star 3 and beyond, programs must both meet a threshold level for the ERS and verify that they meet a sufficient number of points-based quality standards in four domains: Family and Community Partnerships, Qualifications and Professional Development, Management and Administration, and Learning Environment and Curriculum.[3] Details about this joint set of requirements are listed in Table 1.3 as they apply to small and large FCCs, as well as centers.

Notably, within Delaware Stars, providers can choose from an array of standards they wish to be rated on, with more points being required to attain the higher star levels. (However, as described in the next section, Delaware Stars started to phase in a set of required standards [called *essential standards*] to achieve the Star 4 or Star 5 level starting in January 2015.)

Table 1.3. Features of Delaware Stars That Vary by Provider Type as of July 1, 2014

Feature	Small FCCs	Large FCCs	Centers
Eligibility	Licensed providers in good standing	Licensed providers in good standing	Licensed centers in good standing; license-exempt centers in public schools
ERS	FCCERS-R	FCCERS-R	ITERS-R and/or ECERS-R for one-third of randomly selected classrooms
Total standards-assigned points	30	46	46
Total standards, by domain			
Family and Community Partnerships	12 standards (25 points)	13 standards (25 points)	12 standards (20 points)
Qualifications and Professional Development	5 standards (25 points)	14 standards (25 points)	11 standards (30 points)
Management and Administration	4 standards (15 points)	10 standards (20 points)	10 standards (20 points)
Learning Environment and Curriculum	9 standards (35 points)	9 standards (30 points)	13 standards (30 points)
Financial incentives			
QI Grants	up to $750	up to $1,000	By enrollment: 13–60: up to $2,500 61–120: up to $3,000 121–200: up to $4,000 201–300: up to $5,000

SOURCE: Delaware Institute for Excellence in Early Childhood, undated.
NOTES: Features in effect as of July 2014, including the removal of Personal Care Routines from the ERS. Only financial incentives that vary by provider type are listed.

[3] Note that the ERS assessments listed in Table 1.2 differ by program type and age level: Home-based providers are assessed using the Family Child Care Environment Rating Scale–Revised (FCCERS-R; Harms, Cryer, and Clifford, 2007), and center-based care providers that serve infants and toddlers use the Infant/Toddler Environment Rating Scale–Revised (ITERS-R; Harms, Cryer, and Clifford, 2006), whereas centers that serve preschool-age children use the Early Childhood Environmental Rating Scale–Revised (ECERS-R; Harms, Clifford, and Cryer, 2005).

Although the points-based standards are differentiated by program type (small FCCs, large FCCs, and centers), all provider types are required to achieve the same number of total points summed across the four quality domains and minimum ERS scores to advance to Star 3, Star 4, or Star 5 (see Table 1.2).

Several types of programs are eligible to be rated through an alternative pathway: public school programs for preschool-age children with disabilities (called Part B, Section 619 programs under the Individuals with Disabilities Education Act [IDEA]), stand-alone Head Start programs and ECAPs, and programs accredited by the National Association for the Education of Young Children (NAEYC). As shown in Table 1.2, Part B, Section 619 programs enter at Star 2 and move to higher levels based on their ERS score and a review of the standards they meet given their adherence to program standards under their federal funding mechanism. Stand-alone Head Start and ECAP programs can enter the rating system at Star 4 and advance by meeting the ERS threshold only. NAEYC-accredited programs enter at Star 5.

The quality improvement component of Delaware Stars is achieved through general and specialized technical assistance (TA) from trained providers. Once programs join Delaware Stars, they are assigned a TA provider who conducts an onboarding visit and assists providers with creating a QIP. After programs move up to Star 2, TA providers are expected to conduct visits approximately twice a month for those programs actively working to move up a star level and work with the provider to schedule its ERS assessment. Twice-per-month visits continue until the program moves to Star 5, at which point visits are expected to taper off (e.g., one visit per quarter) and only occur in preparation for recertification. Stars Plus, a cohort-based program open to Delaware Stars providers with a large proportion of children with POC subsidies (at least 40 percent of enrolled children), offers more-intensive TA and other supports. Financial incentives are also available to Delaware Stars programs, depending on their rating tier, to support quality improvement and to help programs maintain their quality.

Changes Made to Delaware Stars in 2014–2016

Effective July 2014 (but not implemented until October 2014), Delaware Stars began to phase in a number of enhancements, with additional changes to be phased in through 2015 and 2016 that are discussed later (Delaware Institute for Excellence in Early Childhood, 2014a, 2014b). The initial changes, summarized in Table 1.4, pertain to the number, wording, and points for the points-based standards; the ERS cut scores required for star levels; and the financial incentives available. Although these changes were in effect during our study period, a majority of programs in the validation study were not yet verified under the enhanced rating system, meaning that the study did not reflect the effects of the fully implemented revised Delaware Stars rating structure.

The most significant changes in the structure of the standards occurred in the Learning Environment and Curriculum domain, where the revised standards place greater emphasis on child developmental screenings; observations of children's progress; the use of formative child assessments; implementation of a written comprehensive curriculum aligned with the state's

early learning standards; implementation of a supplemental curriculum to support literacy, mathematics learning, development of social-emotional skills, or healthy lifestyles; use of observations and formative assessments, together with the curriculum, to inform teaching and learning; and support for children with diverse needs. In the case of small and large FCCs, five points were shifted from the Qualifications and Professional Development domain to the Family and Community Partnership domain.

Also effective July 2014, Delaware removed the Personal Care Routines subscale (which pertains to such practices as hand-washing and diapering) from the ERS. The Personal Care Routines subscale had consistently been the lowest-rated domain for Delaware Stars providers and was viewed by OEL as unnecessarily distracting providers' focus from improving the quality of teaching and learning. The items in the Personal Care Routines subscale were also duplicative of OCCL regulations and created confusion for ECE programs. Given the expected higher ERS average scores once this domain was eliminated, the ERS cut score for Star 3 to Star 5 was increased by 0.4 points at each star level.

Finally, with respect to financial incentives, as of July 2014, Delaware eliminated the merit award—a one-time small financial incentive ($300 to $500) for each star level achieved past Star

Table 1.4. Key Enhancements to the Delaware Stars Rating System in 2014

Enhancement	Family Child Care	Centers	Effective Date
Number, wording, and points for points-based standards	• Reduced the number of points-based standards by four standards, two standards, and three standards for small FCCs, large FCCs, and centers, respectively • Revised the wording of some standards • Changed the number of points assigned to some standards		July 1, 2014[a]
Distribution of total points across domains	• Shifted five points from the Qualifications and Professional Development domain to the Family and Community Partnership and Learning Environment and Curriculum domains	No changes made	July 1, 2014[a]
ERS subscales and cut scores	• Removed the Personal Care Routine subscale of the ERS • Increased cut points for ERS to 3.4 for Star 3, 4.4 for Star 4, and 5.4 for Star 5		July 1, 2014
Financial incentives	• Increased tiered reimbursement amounts for Star 4 and Star 5 programs serving infants, toddlers, and preschool-age children • Linked QI Grant amounts to enrollment • Added financial incentives for infant enrollment in Star 4 and Star 5 programs and for programs at the Star 3 through Star 5 levels that were working toward meeting the comprehensive curriculum standard • Eliminated merit awards		July 1, 2014

SOURCES: Delaware Institute for Excellence in Early Childhood, 2014a, 2014b.
[a] According to a personal communication from the Delaware Institute for Excellence in Early Childhood, the changes in the number, wording, and points for the standards were published as effective on July 1, 2014, but they did not go fully into effect until the beginning of October 2014. No programs that were assessed prior to October 2014 were verified using the new version of the points-based standards.

3—in favor of increased funding for POC Tiered Reimbursement Bonuses for Star 4 and Star 5 programs and for program QI Grants that are now tied to program enrollment. Two additional financial incentives were first distributed in December 2014. These incentives include an augmentation to the POC reimbursement, available only for programs at Star 4 and Star 5, based on a program's total infant enrollment. The infant enrollment incentive was added because of the known shortage of high-quality care for infants. A curriculum incentive was also available to programs at star levels 3, 4, and 5 for programs working on the comprehensive curriculum standard. Table A.2 in Appendix A provides more detail on the complete set of financial incentives available to Delaware Stars providers.

The most significant change to the Delaware Stars rating system—a shift from a complete-choice model to a partial-choice model for programs at the two highest star levels—has been phased in over two years, starting in 2015. All programs that submitted requests for assessment and verification after January 1, 2015, were required to meet several "essential standards" to reach (or reverify at) the Star 4 and Star 5 levels. The use of essential standards ensured that top-rated programs would all meet a set of quality standards that were expected to be most important for program quality and promoting child development.

Starting in January 2015, the following four essential standards related to the Learning Environment and Curriculum domain were phased in (at the dates indicated) for both centers and FCC providers that were newly verified or being reverified:

- **Developmental screening.** The program screens each child with a developmental screening tool (such as Ages and Stages) (effective January 2015 for Star 4 and Star 5).
- **Formative assessment.** The program administers a formative child assessment to each child at least twice per year (effective January 2016 for Star 4 and Star 5).
- **Comprehensive curriculum.** The program uses a written comprehensive curriculum that is aligned to the Delaware Early Learning Foundations (effective July 2016 for Star 4 and Star 5).
- **Informing activities, lesson plans, and individualized instruction.** The program uses information from observations, curriculum, formative assessment, Delaware Early Learning Foundations, and families to design daily activities, lesson plans, and individualized goal plans (effective July 2016 for Star 5 only).

In addition, two other essential standards in the Qualifications and Professional Development domain were phased in for center-based providers in order to qualify for Star 4 or Star 5:

- **Administrator credentials.** The administrator completes the Delaware Administrator Credential and qualifies through Delaware First as an administrator (effective July 2015).
- **Curriculum coordinator credential.** The person serving as curriculum coordinator must have the Curriculum and Assessment Credential (effective July 2016).

During the phase-in period, Delaware Stars has been offering a variety of supports to programs to assist them in meeting these required standards (Delaware Institute for Excellence in Early Childhood, 2014b).

Delaware's QRIS in Context

Efforts to design and implement ECE QRISs have been under way in the United States for more than a decade (Zellman and Perlman, 2008). As of May 2016, according to the QRIS National Learning Network (2016), all 50 U.S. states and Washington, D.C., either had a statewide QRIS in place (40 states and Washington, D.C.), had one or more substate QRISs in operation (three states), had a QRIS in the pilot stage (one state), or had a QRIS in the planning stage (seven states). To place Delaware's system in context, we briefly compare Delaware with other states in terms of its child care licensing system—typically the foundation for a QRIS—and the features of its QRIS.

Child Care Licensing

A QRIS is separate from and in addition to a child care licensing system. State licensing systems typically cover both center-based programs and FCC providers of a minimum size and facility type (e.g., school-based early learning programs are often not covered by state licensing systems) (National Center for Early Childhood Quality Assurance, 2015). Licensing standards set requirements that programs must meet—often minimally acceptable levels of program structural features, such as child-staff ratios, square footage per child, and other features relevant for health and safety. Staff criminal background checks and minimum staff age, education, and ongoing training requirements are usually specified as well. Standards are enforced through regular inspections (often unannounced), and more states over time are making inspection reports publicly available on the Internet (National Center for Early Childhood Quality Assurance, 2015).

To be eligible to apply for Delaware Stars, ECE programs must have been licensed for at least six months and be in good standing.[4] Delaware's child care licensing system is rated seventh among the 50 states for the stringency of its requirements and for oversight for licensed child care programs, according to Child Care Aware's most recent ranking from 2013 (Child Care Aware, 2013). The average state obtained an average score of 92 out of 150 possible points. Delaware, by comparison, received 108 points. The top-scoring state, New York, received 116 points. Delaware scored relatively high because of the program standards built into licensing, including requiring child care centers to plan learning activities that address all of the recommended areas, such as active play, early academic development, and social development. Other desired program standards included in Delaware's program are requirements for health

[4] Public schools are an exception; they need not be licensed.

practices in ten specific areas, for safety practices in ten areas, and for programs to encourage parental involvement.

Structure of QRISs

The *QRIS Compendium* (BUILD Initiative, 2016) provides information on the features of the QRISs for systems in operation at the state or local level. Table 1.5 provides a summary of key QRIS features for 37 statewide QRISs as of June 2016.[5] For each QRIS characteristic tabulated, the one that applies to Delaware Stars is shown in bold. This comparison reveals that Delaware Stars, in large part, has adopted a system that shares many features in common with those in other states. With one exception, of the features listed in Table 1.5, Delaware Stars has the most commonly selected configuration among the statewide QRISs, such as a voluntary system (31 QRISs), a rating scale with five tiers (23 QRISs), tiered reimbursement (26 QRISs), reliance on the Child Care Development Fund or an RTT-ELC grant as a funding source (27 states and 13 states, respectively), and use of the ERS family of observation tools as part of the rating determination (26 states). While 16 states use a block system in their rating structure, Delaware is one of 15 states that uses a hybrid system, combining a block structure with a point-based system.[6]

QRIS Evaluations

QRISs combine multiple indicators of ECE program quality into a single summary rating to make program quality more transparent to consumers and funders in the public and private sectors. As QRISs have matured, states and localities have undertaken evaluations to answer a number of questions about QRIS design and validity (Zellman and Fiene, 2012; Karoly, 2014; Lahti et al., 2015). Indeed, according to the *QRIS Compendium*, all but five of the 37 QRISs summarized in Table 1.5 have one or more evaluation studies under way or completed. The prevalence of evaluations has expanded, in part, because of the requirement under the RTT-ELC grant mechanism to conduct an independent evaluation of the state's QRIS. Ideally, such evaluations are conducted through time as QRISs are designed, piloted, and brought to scale, with opportunities for continuous improvement of the QRIS based on evaluation findings at each stage (Zellman and Fiene, 2012).

---

[5] Following the *QRIS Compendium* (BUILD Initiative, 2016), we include a single QRIS for California, even though it is being implemented at the county level, because it is operating in over half the counties and follows a common structure, albeit with some local options. We do not include the three distinct county-level QRISs operating in Florida.

[6] In a block structure, programs must meet all quality standards at a given quality tier to be rated at that level. Under a points-based system, points are associated with achieving specific quality standards, and the rating tier is based on total points achieved. A hybrid structure combines the two approaches, with rating tiers based on meeting both designated quality standards and a specified standards-based point total.

Table 1.5. Characteristics of 37 Statewide QRISs

QRIS Characteristic	Specific Feature	Number of States	States
Mandate for ECE programs to participate in the QRIS	Mandatory for all licensed programs	1	CO
	Mandatory depending on funding or program type	5	AR, KY, RI, WA, WI
	Voluntary	**31**	AZ, CA, **DE**, GA, IA, ID, IL, IN, MA, MD, ME, MI, MN, MS, MT, NC, ND, NE, NH, NM, NV, NY, OH, OK, OR, PA, SC, TN, UT, VA, VT
Number of rating tiers (e.g., number of star levels)	6 tiers	1	ID
	5 tiers	**23**	AZ, CA, CO, **DE**, IA, IL, MD, MI, MS, MT, NC, NE, NV, NY, OH, OR, RI, SC, UT, VA, VT, WA, WI
	4 tiers	9	IN, KY, MA, ME, MN, ND, NM, OK, PA
	3 tiers	4	AR, GA, NH, TN
Rating structure	Block	16	AR, ID, IL, IN, KY, MA, MD, ME, MS, MT, ND, OK, OR, PA, RI, VA
	Points	6	GA, MI, NC, NY, UT, VT
	Hybrid	**15**	AZ, CA, CO, **DE**, IA, MN, NE, NH, NM, NV, OH, SC, TN, WA, WI
Length of time rating is valid	1 year	9	ID, IN, MD, MS, MT, NM, PA, UT, WI
	18 months	1	NV
	2 years	6	CA, IA, MA, MI, MN
	3 years	**14**	AR, CO, **DE**, GA, IL, ME, NC, ND, NH, NY, OR, RI, TN, VT, WA
	Varies by rating level	6	AZ, KY, NE, OH, SC, VA
	Nonexpiring	1	OK
Reimbursement rates are tiered by rating level	**Yes**	**26**	AR, CO, **DE**, GA, IL, IN, MA, MD, ME, MI, MN, MT, NC, NE, NH, NV, NM, OH, OK, OR, PA, SC, TN, VT, WA, WI
	No	11	AZ, CA, IA, ID, KY, MS, ND, NY, RI, UT, VA
Funding sources for QRIS (according to those states that completed self-reports; more than one may apply)	**Child Care and Development Fund**	**27**	AR, CO, **DE**, GA, IA, ID, IN, ME, MI, MN, MT, NC, NH, NM, NV, NY, OK, OR, PA, RI, SC, TN, UT, VA, VT, WA, WI
	Philanthropic funds	8	GA, ME, MN, NE, NM, NY, NV, WA
	State education funds	**5**	CA, **DE**, IL, MS, NC
	State human services funds	1	OK
	State general funds	9	CO, IA, MI, ND, NM, NY, OR, PA, VT
	State preschool funds	**8**	AR, AZ, **DE**, IL, MA, NC, NM, OR
	RTT-ELC	**13**	CA, CO, **DE**, GA, IL, MA, MD, MI, MN, OR, RI, VT, WA
	TANF	3	IA, OK, WI
Observational tools used for ratings	CLASS	16	AZ, CA, CO, IL, MA, MD, MN, ND, NE, NY, OK, OR, RI, VA, VT, WA
	ERS	**26**	AR, AZ, CA, CO, **DE**, GA, IA, ID, IL, KY, MA, MD, MS, MT, NC, ND, NM, NV, NY, OK, PA, SC, TN, VA, VT, WI
	PQA	2	AR, MI
	TPOT and TPITOS	1	MT
	Self-designed	2	OH, SC
	Optional	1	NH
	No observational tools used	3	IN, ME, UT

SOURCE: BUILD Initiative, 2016.
NOTES: QRIS features are as of June 2016. State postal abbreviations are listed alphabetically. The features of Delaware Stars are shown in bold. TANF = Temporary Assistance for Needy Families. TPITOS = Teaching Pyramid Infant Toddler Observation Scale. TPOT = Teaching Pyramid Observation Tool.

A central question that has been part of many of these QRIS evaluations is whether the rating system is valid—i.e., does it capture meaningful differences in program quality in moving up the rating scale? Validation of a QRIS has typically been addressed in two ways.[7] One approach is to determine whether programs with higher ratings in the QRIS have higher quality, based on global or specific measures of quality that are not used in the rating system itself.[8] A second approach is to evaluate whether children in programs with higher quality ratings demonstrate better developmental outcomes (or larger developmental gains) when compared with children in lower-rated programs. Again, according to the *QRIS Compendium*, either or both of these approaches have been incorporated as validation studies into 21 of the 37 QRISs included in Table 1.5. Here we highlight key findings from this body of research, drawing on the available published studies, particularly those for RTT-ELC states.

QRIS Ratings and Alternative Measures of Quality

Table 1.6 provides a summary of the first type of validation study, in which the relationships between QRIS ratings and other validated measures of ECE program quality were examined. With a few exceptions (e.g., the Colorado study by Zellman et al., 2008), these eight studies found that at least some alternative measures of program quality were positively (and usually significantly) correlated with QRIS ratings. At the same time, these studies generally found that these relationships were weak, such that differences in quality were significant only at the extremes—i.e., the highest tier versus the lowest tier—and even then, the differences in quality were not especially large when moving up the ratings tiers.

Two of the earliest QRIS validation studies using the ERS as the alternative quality measure—one for Indiana and the other for Maine—illustrate this point (Lahti et al., 2015). The validation study of Indiana's QRIS found that, for centers and FCC providers combined, the mean ERS score at tier four of 4.3 was significantly different than the mean ERS score at tier one of 3.2 (Elicker et al., 2011). But that differential on the seven-point ERS rating scale of 1.1 scale points was not a particularly large difference in quality when moving from the lowest rating level to the highest, especially considering that two scale points differentiate programs in moving from quality that is "inadequate" (a score of 1), to "minimally acceptable" (a score of 3), to "good" (a score of 5), to "excellent" (a score of 7). For center-based programs, the differential in the mean ERS score from tier one to tier four was even smaller: just 0.5 scale points. Equally modest differentials were found for the Arnett Caregiver Interaction Scale (CIS), a measure of the quality of teacher-child interactions (Arnett, 1989). Likewise, the study for Maine's QRIS

---

[7] Zellman and Fiene (2012) identify two other approaches to QRIS validation, often taken as initial steps in the design and piloting of a rating scale: (1) examine the validity of the underlying concepts in the rating scale and (2) assess the measures used in the rating scale and their psychometric properties.

[8] Some early QRIS validation studies examined the relationship between program ratings and measures of quality that were part of the rating system, a weaker validation design because of the inherent correlation between the quality measure and the quality rating tier. See Karoly (2014) for a summary of those studies.

Table 1.6. Evaluations of QRIS Ratings and Program Quality

Study/QRIS	Settings/Sample	Alternative Measure of Quality	Key Findings
Zellman et al. (2008) Colorado Qualistar	• 65 centers (Wave 1) • 38 FCCs (Wave 1)	• CIS • Pre-K Snapshot subscales	• QRIS ratings for centers were significantly positively related to two of the four CIS subscales (detachment and positive relationship) but not to any of the Pre-K Snapshot subscales (Wave 1 data only). • QRIS ratings for FCCHs were not significantly related to the CIS or the Pre-K Snapshot subscales (Wave 1 data only).
Elicker et al. (2011) Indiana Paths to Quality (PTQ)	• 135 classrooms in 95 licensed centers • 169 licensed FCCs • 12 unlicensed registered child care ministries	• ERS (ITERS-R, ECERS-R, FCCERS-R) • CIS	• QRIS ratings were significantly positively associated with CIS and ERS scores—as scores increased, so did ratings. • ERS scores were highly variable within each rating level for all QRIS levels and all types of care.
Lahti et al. (2011) Maine Quality for ME	• 194 classrooms in 142 centers • 113 FCCs	• ERS (ITERS-R, ECERS-R, SACERS, FCCERS-R)	• QRIS ratings were significantly positively correlated with ERS.
Tout et al. (2016) Minnesota Parent Aware	• 146 centers (ERS); 261 centers (CLASS) • 57 FCCs	• ERS (ECERS-R, ECERS-E, FCCERS-R) • CLASS (centers only)	• On four of seven observed quality measures for centers, quality was significantly higher for Star 3 and Star 4 programs when compared with Star 1 and Star 2 programs; there were no significant differences for CLASS. • There were no significant differences for FCC providers in the quality measures across rating tiers.
Norris, Dunn, and Eckert (2003) Oklahoma Reaching for the Stars	• 336 centers with at least one preschool room • Assessments for 279 infant/toddler rooms, 336 preschool rooms, and 152 school-age rooms	• ERS (ITERS, ECERS-R, SACERS) • CIS	• Classroom ERS scores improved with each rating tier (4 total), with statistically significant differences in all pairwise tier comparisons made for ECERS-R and in 4 of 5 comparisons made for ITERS and SACERS. • Classroom CIS scores improved with each rating tier, but differences were statistically significant only for infant/toddler rooms.
Norris and Dunn (2004) Oklahoma Reaching for the Stars	• 189 FCCs	• CIS	• Two-Star FCC providers were more sensitive in their interactions with children than One-Star providers were, as measured by the CIS. • Sample sizes were too small to analyze Three-Star (the highest category) providers.
Sirinides (2010) Pennsylvania Keystone STARS	• Sample of 88 classrooms in STAR 3 or STAR 4 centers for CLASS administration	• CLASS	• Scores were higher for STAR 4 classrooms compared with STAR 3 classrooms on all CLASS subscales.
Magnuson and Lin (2015) Wisconsin YoungStar	• 204 classrooms in 120 centers • 35 classrooms in FCCs	• ERS (ECERS-R, FCCERS-R)	• ERS scores were significantly higher for Star 2 as compared with Star 3 through Star 5.

SOURCE: Cited studies.

showed statistically significant differences in mean ERS scores between the lowest- and highest-quality tiers in the four-tier system, but again the differentials were 0.5 scale points for centers and 0.9 scale points for FCC providers on the seven-point ERS scale (Lahti et al., 2011). Moreover, at the time of both of these validation studies, providers at the highest rating tiers had mean ERS scores below the "good" to "excellent" quality range (a score of 5 or higher on the ERS).

This basic pattern has been replicated in the most recent QRIS validation studies conducted as part of RTT-ELC grants. For example, the validation study for Wisconsin's QRIS demonstrated significant differences in ERS levels between Star 2 and the average of Star 3 to Star 5, but again the differential from the lowest to highest levels was modest (0.8 scale points for centers). Likewise, the average ERS score at the highest level (4.6 for the ECCERS-R) did not reach the "good" to "excellent" range (Magnuson and Lin, 2015).

A similar result was found for the most recent validation study of Minnesota's QRIS, where the ERS scores were significantly different between low- and high-rated programs, but the differential was just 0.3 scale points between Star 1 and Star 2 programs on average, compared with Star 3 and Star 4 programs on average (Tout et al., 2016).[9] Again, average quality for programs rated in the top two tiers was 4.0 on the ERS, one scale point below the "good" level of quality. There were no significant differences across rating tiers based on the most commonly used measure of teacher-child interactions, CLASS (Pianta, La Paro, and Hamre, 2008) (despite being a tool that is used in the rating scale to reach the top two tiers). On average, scores on the seven-point CLASS subscales for Classroom Organization and Emotional Support were 6.1 and 6.2 to 6.3 for lower- and higher-rated programs, respectively, which are in the high-quality range. Average scores for the CLASS Instructional Support subscale were 2.4 to 2.5 for lower- and higher-rated programs, respectively, below the level attained for effective programs.[10]

QRIS Ratings and Child Developmental Outcomes

Fewer validation studies to date have examined the association between QRIS ratings and children's developmental outcomes. As discussed in Karoly (2014), to assess whether ECE program ratings are predictive of child development, the ideal study design would randomly assign children to programs of varying levels of rated quality (i.e., Star 1 programs, Star 2 programs, etc.) at the start of the program year and then measure domains of child development after a sufficient time had passed. If the program ratings captured meaningful differences in

[9] See also Tout et al. (2010, 2011) for earlier rounds of data collection, when Minnesota's QRIS was being piloted, using similar methods that produced similar findings.

[10] The Instructional Support domain consistently receives the lowest scores of the three CLASS subscales, and even programs that have demonstrated favorable impacts on school readiness outcomes in rigorous evaluations score relatively low in this domain. For example, classrooms in the highly effective Tulsa, Oklahoma, universal preschool program for four-year-olds received an average score of 3.2 on the seven-point scale (Gormley et al., 2005), and classrooms in Boston's proven preschool program for four-year-olds were assessed with an average Instructional Support score of 4.3 (Weiland and Yoshikawa, 2013; Weiland et al., 2013).

quality, we would expect to see higher levels of child development, on average, for programs at each successive rating tier. In the absence of such random assignment, if we simply examine the average level of child development for programs in different rating tiers at a point in time, it is possible that the differences we observe are at least in part the result of selectivity—i.e., parents choose to enroll their children in programs based on the program characteristics, including dimensions of quality. If parents with more resources choose higher-quality programs, then the higher levels of development for the children in those programs is likely attributable to some combination of family background factors and the impact of the program itself. Without accounting for the impact of selectivity when using observational data, estimates of the relationship between program quality and child outcomes will be biased.

Because no studies to date have had the option of random assignment, researchers have employed various research designs to try to mitigate the potential for selectivity bias. One option is to include child and family background characteristics, measured through a parent survey or through administrative program data (e.g., eligibility for free or reduced-price lunch), to control for observed factors that may influence the selectivity of children into ECE programs. A second option is to use a pretest–posttest design, where gains in child development are calculated from a baseline (pretest) to a follow-up wave (posttest). This approach controls for differential levels of development at the baseline. Controls for family background in such a longitudinal analysis may further diminish any selectivity bias. We view this as the preferred study design, which is the approach taken in the five published QRIS validation studies summarized in Table 1.7.[11]

Taken together, the studies in Table 1.7 provide only limited evidence that programs rated more highly in a given QRIS are associated with better developmental outcomes for the enrolled children. Three of the five studies found a relationship between QRIS ratings and children's development for at least one of the developmental domains assessed. Most recently, the validation study for Minnesota's QRIS found significant relationships between quality rating and children's development for just two of the nine developmental assessments, both social-emotional measures. The study of Missouri's QRIS, conducted during the pilot phase, reported that more highly rated programs were associated with better developmental outcomes, although only for a limited set of the social-emotional measures (Thornburg et al., 2009). Notably, none of the nine measures of early reading or quantitative skills showed a relationship with rated program quality for the full sample. When the analysis was stratified by child poverty status, a measure of receptive vocabulary (the Peabody Picture Vocabulary Test–4 [PPVT–4]) showed the expected relationship with the quality tier for the subset of children in poverty.

The third longitudinal study for Virginia's QRIS—while relying solely on a teacher-performed assessment of preliteracy skills in two areas (Alphabet Knowledge and Phonological Awareness) for the set of state-funded prekindergarten programs in the QRIS—included a potentially richer set of control variables measured at the child, center, and community levels.

[11] Findings for studies using other measures are summarized in Karoly (2014).

Table 1.7. Evaluations of QRIS Ratings and Child Developmental Outcomes Using Longitudinal Designs

Study/QRIS	Settings/Sample	Methods	Measures of Child Development	Key Findings
Zellman et al. (2008) Colorado Qualistar	1,368 preschool-age children enrolled in QRIS-rated centers or FCCs in Wave 1; 829 children in Wave 2; 619 children in Wave 3	• Longitudinal (3 points in time) • Family background controls (parent survey) • Primary data	Direct assessment • PPVT–4 • WJ–III Letter Word Identification • WJ–III Passage Comprehension • WJ–III Applied Problems Teacher assessment • Child Behavior Inventory (CBI) Parent assessment • Strength and Difficulties Questionnaire (SDQ) (Wave 3 only)	• QRIS ratings were not associated with improvement in child outcomes for either centers or FCCs. • Individual components of the QRIS ratings (e.g., average class ratio, parent survey, head teacher educational attainment) were not associated with any improvement in child outcomes. • Subgroup analyses did not show that low-income children were more likely to benefit from highly rated centers.
Tout et al. (2016) Minnesota Parent Aware	1,181 preschool-age children enrolled in 325 QRIS-rated centers or FCCs	• Longitudinal (fall to spring) • Child level • Family background controls (parent survey) • Primary data	Direct assessment • Test of Preschool Early Literacy (TOPEL) Phonological Awareness and Print Knowledge • WJ–III Applied Problems • Bracken School Readiness Assessment • Peg Tapping Test • Height and weight Teacher assessment • Social Competence and Behavior Evaluation short form (SCBE-30) • Preschool Learning and Behavior Scale (PLBS) Attention/Persistence subscale	• Significant differences across rating tiers (Star 1 and Star 2 versus Star 3 and Star 4) in developmental gains were found for two of nine measures of development (social competence and persistence). • Low-income children in higher-rated programs had great gains on two of nine measures (print knowledge and social competence).

Table 1.7. Evaluations of QRIS Ratings and Child Developmental Outcomes Using Longitudinal Designs, Continued

Study/QRIS	Settings/Sample	Methods	Measures of Child Development	Key Findings
Thornburg et al. (2009) Missouri Pilot Quality Rating System	350 preschool-age children in 66 classrooms enrolled full time (25+ hours) in 32 licensed centers and 6 licensed FCCs (excluded non-English speakers and those with severe disabilities)	• Longitudinal (fall to spring) • Child level • Family background controls (parent survey) • Primary data	Direct assessment • PPVT–4 • TERA–3 Reading Quotient • TERA–3 Alphabet subtest • TERA–3 Conventions subtest • TERA–3 Meaning subtest • WJ–III Applied Problems • Shape identification • Color identification • Upper-case alphabet • Fine motor skills • Gross motor skills • DECA Total Protective Factors • DECA Initiative scale • DECA Self-Control scale • DECA Attachment scale • DECA Behavioral Concerns	For all children by rating tier, statistically significant greater gains were found for the following outcomes (effect sizes in parentheses): • high (4–5 stars) versus low (1–2 stars): overall social and behavioral skills (0.80), motivation (0.79), self-control (0.65), and positive adult relationships (0.45) • medium (3 stars) versus low (1–2 stars): overall social and behavioral skills (0.36) and motivation (0.43). For children in poverty by rating tier, statistically significant greater gains were found for the following: • high versus low: overall social and behavioral skills (0.79), motivation (0.78), and vocabulary (0.74) • medium versus low: vocabulary (0.64) • high versus medium: self-control (0.61). For children not in poverty by rating tier, statistically significant greater gains were found for the following: • high versus low: overall social and behavioral skills (0.79), motivation (0.79), and self-control (0.66) • medium versus low: overall social and behavioral skills (0.49), motivation (0.57), and positive adult relationships (0.33).

Table 1.7. Evaluations of QRIS Ratings and Child Developmental Outcomes Using Longitudinal Designs, Continued

Study/QRIS	Settings/Sample	Methods	Measures of Child Development	Key Findings
Sabol and Pianta (2012) Virginia Star Quality Initiative	2,805 preschool-age children in 71 QRIS-rated state-funded prekindergarten programs	• Longitudinal (fall to spring in pre-K and K year) • Child level • Family background controls (child record) • Center and community characteristics controls (or fixed effects) • Primary data	Pre-K and K teacher assessment of preliteracy skills • Phonological Awareness Literacy Screening (PALS) Pre-K (seven subtests; used to derive two factors: Alphabet Knowledge and Phonological Awareness) • Phonological Awareness Literacy Screening (PALS) K (seven subtests; used to derive two factors: Alphabet Knowledge and Phonological Awareness)	• There was no correlation between pre-K star levels and fall K preliteracy skills after controlling for pre-K fall preliteracy skills, family background, center characteristics, and community characteristics. • Using the same controls, the growth in Alphabet Knowledge during the pre-K year was significantly higher for children in 3-star programs versus 2-star programs (effect size of 0.43) and in 4-star programs versus 2-star programs (0.40); the growth in Phonological Awareness in the pre-K year was significantly higher only for children in 3-star programs versus 2-star programs (0.37). • Using the same controls, compared with 2-star programs, children in 3-star and 4-star programs had significantly higher declines in Alphabet Knowledge between the spring pre-K and fall K assessments (effect sizes of –0.12 and –0.18, respectively). • Using the same controls, there was no difference in fall-spring growth during the K year by pre-K star rating.
Magnuson and Lin (2016) Wisconsin YoungStar	725 preschool-age children in 151 QRIS-rated centers or FCCs	• Longitudinal (fall to spring) • Child level • Family background controls (parent survey) • Primary data	Direct assessment • WJ–III Letter Word Identification • WJ–III Applied Problems • Bracken School Readiness Assessment—Third Edition • TOPEL Phonological Awareness • Head-Toes-Knees-Shoulders (HTKS) Teacher assessment • SCBE-30 Social Competence, Anger-Aggression, and Anxiety-Withdrawal • PLBS	• There were no significant differences in child developmental gains for children in Star 2 programs versus Star 3 to Star 5 programs on any of the direct assessments or teacher assessments.

SOURCE: Cited studies.
NOTE: TERA–3 = Test of Early Reading Ability, Third Edition.

This study of Virginia's QRIS demonstrated significantly higher gains during the prekindergarten year for four-star versus two-star programs and three-star versus two-star programs for one or both of the preliteracy measures. At the same time, there was no indication that program quality as rated by the QRIS was associated with subsequent performance on the literacy measures during the kindergarten year.

In contrast, one of the first validation studies—for Colorado's Qualistar QRIS (Zellman et al., 2008)—and one of the most recent RTT-ELC validation studies—for Wisconsin's YoungStar QRIS (Magnuson and Lin, 2016)—both showed that, as configured, the rating systems did not generate quality ratings that distinguished programs in terms of the developmental gains experienced by participating children. Both of those studies included both early academic and noncognitive measures of child development.

Limitations

It is important to acknowledge several limitations of our study that have implications for our findings and the conclusions we draw. We discuss these issues in the context of our analyses and findings, particularly in Chapters 2, 3, and 4, but it is useful to highlight them at the outset.

First, as discussed earlier in this chapter, Delaware Stars has undergone several structural changes since its inception in 2007. Of importance for our analysis is that the changes to the rating structure that became effective starting in July 2014 were not fully implemented prior to conducting our evaluation. During the period covered by our data collection (fall 2014 to spring 2015), most programs in Delaware Stars had yet to have their ratings determined using the new ratings structure—particularly the use of essential standards, which were not fully phased in until July 2016. This means that we have not been able to assess the validity of the fully phased-in rating structure. Thus, this evaluation should be viewed as a baseline for Delaware Stars, against which future Delaware Stars validation studies can be compared.

Second, given the cost of collecting primary data on program quality and children's development for all licensed programs in Delaware, our study design, like those of other states' QRIS validation studies, collected information for a sample of providers. Given the response rates that we detail in Chapter Two, we are limited in our ability to assess the validity of the Delaware Stars ratings for FCC providers. Our findings are most relevant for school- and center-based providers, which are the ECE settings attended by most children in Delaware.[12] In addition, by design, our study did not examine care quality or developmental outcomes for infants in center- or home-based care settings.

Third, as with other validation studies, our analyses of the relationship between program quality and children's development are potentially affected by the role that parental choice plays

[12] For example, as of October 2014, over 70 percent of center-based providers participated in Delaware Stars, compared with about 25 percent of small and large FCCs (Auger, Karoly, and Schwartz, 2015). Overall, about 80 percent of children enrolled in licensed ECE settings were in center-based programs as of that date.

in which programs children attend.[13] (Another element of choice is which parents provided consent for us to conduct developmental assessments of their children.) Although our empirical models control for a number of relevant child and family background characteristics, we cannot rule out possible selectivity bias. At the same time, if we expect children from families with greater advantages to attend higher-quality programs, the selection bias would make us likely to overestimate the relationship between program quality and child development because children in higher-quality programs would have better outcomes in part because of their more-advantaged circumstances, rather than only because of the higher program quality. This would make it more likely for us to find a stronger relationship between program quality and child development than actually exists.

Fourth, as with other similar validation studies, our analyses are affected by the limitations of the measures of program quality and child development that we employ. For example, a number of recent studies demonstrate relatively weak associations between the measures of ECE program quality commonly used in QRIS rating systems (e.g., the ERS) and those used in QRIS validation studies (e.g., the CLASS) (Auger et al., 2012; Burchinal, 2010; Burchinal, Kainz, and Cai, 2011; Sabol et al., 2013; Weiland et al., 2013). Likewise, other commonly used measures of program quality, such as the education level of the lead teacher, do not consistently predict which programs have higher quality or better child outcomes (Karoly, 2012). This issue is a general challenge for the field, as the premise for a QRIS is that there are measures of quality that can be used to detect meaningful differences in program quality that can then be incorporated into the rating structure. Further, in terms of child development, while we were able to assess children's development in several areas using multiple assessments with demonstrated reliability and validity, our conclusions apply only for those areas of development that we were able to measure.

Fifth, there are also weaknesses with the administrative data we employ from the Delaware Stars system. We have limited current information on program features for all licensed providers, such as current enrollment or whether the provider accepts children with child care subsidies. These measurement limitations may also affect our ability to find strong relationships in the data.

Despite these concerns, this evaluation is an important step in the process of implementing and sustaining an effective ECE QRIS. The analyses reported in the chapters that follow adopt the most rigorous methods employed in similar QRIS validation studies, using data collected by well-trained assessors and using best practices for data quality assurance. Ultimately, the analyses will provide OEL and other stakeholders with independent, objective, and rigorous empirical evidence of the extent to which rating tiers reflect relevant differences in the quality of

[13] The ideal study design would allow us to randomly assign children to providers, so that whether a child attended a lower- or higher-quality provider was determined by chance. With such an experimental design, we would be more confident that any relationship we saw between program quality and children's developmental outcomes was the result of the program's quality, rather than other unmeasured factors (e.g., child and family background characteristics that we did not observe).

home- and center-based providers and whether the state's QRIS is operating effectively in terms of TA, financial supports, and other features. The study also provides a baseline from which to build future evaluations of Delaware Stars.

Road Map for the Report

We begin in the next chapter with a description of the primary data collected from a sample of ECE providers in Delaware and the children enrolled in their programs. Chapter 3 presents results from our analysis of the relationship between ratings in Delaware Stars and the measures of observed quality collected for the provider sample. We also consider the relationship between components of the rating structure (e.g., the ERS score and the points obtained in the four quality domains) and our alternative measures of program quality. Chapter 4 considers the relationship between Delaware Stars ratings and child learning and development in early academic, cognitive, social-emotional, and behavioral skills. Chapter 5 focuses on the remaining study questions listed in Table 1.1, drawing on findings from our survey of program directors and their experience with Delaware Stars. We also integrate findings from earlier reports. Chapter 6 summarizes our key findings and draws out the implications for Delaware Stars and QRISs more generally. A series of appendixes provide supporting documentation and detailed empirical results.

Chapter 2. Data Sources and Measures

To answer the primary research questions, we collected data from participating ECE programs and from children in those programs whose parents consented to participate in the research study. We interviewed program directors, observed classrooms, and assessed children ages two to five.[14] We describe the provider sample, the data collected, and the resulting sample of providers and children. Additional details about the sampling and recruitment of providers, field staff recruitment and training, data collection methods, and measures of provider quality and children's development are provided in Appendixes B and C. In Appendix D, we provide further detail on the characteristics of the programs in our sample, their directors, and the children who participated in the study.

Provider Sample

To validate Delaware Stars, we sought study participation from a representative group of ECE programs both in and out of Delaware Stars and within each of the three license types: small FCCs, large FCCs, and centers. Such programs include those with public funding through Head Start, POC, and ECAP, as well as those paid for by parents or other private sources. School-based programs in Delaware Stars were included in the sample frame as well.

In May 2014, RAND sent a study invitation packet to all ECE programs in Delaware Stars at the time; all licensed centers and all licensed large FCC programs *not* in Delaware Stars; and a randomly selected sample of licensed small FCCs also not in Delaware Stars. During the recruitment period, to track participation rates, we segregated centers with more than and fewer than 50 enrollees to ensure that we were obtaining equal rates of participation in the study from both types of centers. We obtained refreshed Delaware Stars rating data as of August 2014 that we used as our sampling reference throughout the study. As shown in Table 2.1, out of nearly 1,200 eligible providers, RAND invited a total of 774 providers to participate in the study, and we ultimately collected data from 181 providers, which we refer to as the provider sample. The response rate calculations, detailed in Appendix B, indicate an overall response rate of 25 percent. Response rates were considerably higher for providers in Delaware Stars, the programs of primary interest in our analyses. Response rates reached 48 and 46 percent, respectively, among public school programs and centers in Delaware Stars and 23 percent and 15 percent, respectively, for small and large FCCs in Delaware Stars. Response rates were lower for

[14] For simplicity, we use the term *classroom* in the context of center-based programs but also to refer to the group or room in which children receive care in small and large FCCs. Likewise, we use *program director* to refer to center leaders as well as FCC providers, and we use *teacher* also to mean *caregiver*.

providers not in Delaware Stars—specifically, 23 percent for centers and 1 and 2 percent for small and large FCCs.

Table 2.1 also shows that the Delaware Stars status of programs in the provider sample changed between August 2014, when we drew the sample (third column in Table 2.1), and August 2015, when data collection was completed (final column in Table 2.1). In particular, two small FCCs left Delaware Stars during that period, while one large FCC entered Delaware Stars. Among the center-based programs, ten of the 21 centers originally not in Delaware Stars enrolled, but one center in Delaware Stars left, for a net gain of nine additional centers in Delaware Stars by the end of the field period. Some providers in Delaware Stars also moved up in their ratings between August 2014 and August 2015. As described later, we use Delaware Stars status as of August 2015 in our analyses.

To generalize our sample to all licensed ECE programs as of August 2015, we developed and applied weights using methods described in Appendix B. Notably, the response rate analysis demonstrated that provider type and Delaware Stars status were significant predictors of which providers participated in the study, with patterns consistent with those shown in Table 2.1. There were no differences in response rates based on provider county or community characteristics in the provider's zip code. We applied the provider weights in all analyses of the provider sample reported in Chapters 3 and 5. Because of the low response rate for providers not in Delaware Stars, our weighted sample of providers is most representative of the underlying population of providers when we focus on providers participating in Delaware Stars.

Table 2.1. Provider Response Rates in the Validation Study

Provider Type	Providers as of August 2014				Provider Sample Distribution as of August 2015
	Number in Delaware[a]	Number Invited to Participate in Study	Number in Provider Sample[b]	Response Rate (%)	
In Delaware Stars					
Small FCCs	162	162	35	23.0	33
Large FCCs	26	26	4	15.4	5
Centers	241	241	106	46.1	115
Public schools	25	25	12	48.0	12
Not in Delaware Stars					
Small FCCs	596	174	2	1.2	4
Large FCCs	47	47	1	2.4	0
Centers	99	99	21	23.1	12
All providers	1,196	774	181	24.9	181

SOURCE: Authors' analysis of provider sample disposition file and response rate analysis.
NOTES: Although 214 total programs signed a consent form to participate in the study, 181 ultimately participated in one or more data collection activities. Response rate analysis is provided in Table B.2.
[a] Total includes all licensed small FCCs, large FCCs, and centers as of August 2014, plus public school programs participating in Delaware Stars as of May 2014. It does not include public school programs that were not participating in Delaware Stars.
[b] Consists of providers that participated in one or more data collection activities (i.e., provider survey, child assessments, or classroom quality observations).

Types of Data Collected from Participating Programs and Children

RAND field staff collected data in three time periods that began in October 2014 and ended in August 2015. (See Appendix B for additional detail on the fieldwork.) Specifically, we collected the following data (see Table 2.2 for the counts of cases for which we collected data in each category):

Director Interview

To understand directors' perceptions of Delaware Stars and to collect information about their programs, we interviewed directors in the provider sample once in fall 2014 and a second time in spring 2015. (For small FCCs, we interviewed directors only once, in fall 2014.) The director interview used a structured interview protocol (included in Appendix C), and most interviews were conducted by phone.

Table 2.2. Data RAND Collected for the Validation Study

	Data Collection Period		
Data Component	**Fall 2014 (October 2014–February 2015)**	**Winter 2015 (February 2015–July 2015)**	**Spring 2015 (April 2015–August 2015)**
Director interview			
Small FCCs	21 programs	–	33 programs
Large FCCs	5 programs	–	4 programs
Centers (small and large)	130 programs	–	129 programs
Parent survey (child and family characteristics)			
Large FCCs	18 children (5 programs)	–	–
Centers (small and large)	1,297 children (138 programs)	–	–
Child assessments			
Large FCCs	18 children (5 programs)	–	18 children (5 programs)
Centers (small and large)	1,297 children (138 programs)	–	1,105 children (135 programs)
Classroom quality observation			
Small FCCs	–	35 classrooms (35 programs)	–
Large FCCs	–	5 classrooms (5 programs)	–
Centers (small and large)	–	315 classrooms (132 programs)	–
Program license and Delaware Stars data (e.g., license type, Star rating, and components)	For all licensed programs as of August 2014	–	For all licensed programs as of August 2015
Neighborhood characteristics of ECE programs	For all licensed programs based on their zip code	–	–

NOTES: The number of children assessed includes children who completed at least one, if not all, of the types of assessments administered. Likewise, the number of classrooms reported includes classrooms for which RAND has at least one out of the three quality measures. Although not all 181 programs in the analytic sample completed both the fall and the spring director interview, each completed at least one. – = not applicable.

Parent Survey

Following procedures detailed in Appendix B, parents of age-eligible children enrolled in large FCCs and centers in the provider sample programs were invited to give active consent for their children to participate in two waves of child developmental assessments.[15] Along with consent, we included a short self-administered survey with seven questions about parent or guardians' education, income, household size, home zip code, and language spoken at home. The survey instrument is provided in Appendix C.

Child Assessments

To measure children's learning, well-trained field staff assessed children in the large FCCs and centers in the provider sample for whom we had parental consent, first in fall 2014 and again in spring 2015, using valid and reliable measures, all of which had been used in other QRIS validation studies. The assessments, listed in Table 2.3, included three direct assessments of early academic skills—specifically, vocabulary, early reading skills, and early mathematics skills. A fourth performance-based cognitive assessment measured aspects of executive function (e.g., attention, inhibitory control, and working memory). These four were directly administered to children and took a combined total of 45 minutes on average. A final assessment, completed by teachers, measured both positive and negative aspects of social-emotional and behavioral skills. Children could be assessed in English alone or in English and Spanish, following a protocol described in Appendix C.

Table 2.3. Measures of Child Development

Measurement Tool	What It Measures
Assessments of Early Academic and Cognitive Skills	
Peabody Picture Vocabulary Test (4th edition) (PPVT–4) (Dunn and Dunn, 2007)	Direct assessment of receptive vocabulary (ability to understand spoken words)
Woodcock-Johnson III (WJ–III) Letter Word Identification (Woodcock, McGrew, and Mather, 2006)	Direct assessment of early reading skills (e.g., ability to match a pictographic representation of a word with an actual picture of the object and identifying letters and words)
Woodcock-Johnson III (WJ–III) Applied Problems (Woodcock, McGrew, and Mather, 2006)	Direct assessment of skill in solving practical math problems (e.g., counting, addition, subtraction)
Head-Toes-Knees-Shoulders (HTKS) (Cameron and McClelland, 2011)	Performance-based assessment of executive function (e.g., attention, inhibitory control, and working memory)
Assessment of Social, Emotional, and Behavioral Skills	
Devereaux Early Childhood Assessment (DECA) (Mackrain, LeBuffe, and Powell, 2007; LeBuffe and Naglieri, 2012)	Caregiver/teacher-rated assessment of • protective factors (overall positive social and behavior skills) • behavioral concerns (incidence of emotional and behavioral problems)

NOTE: See Appendix B for citations to the Spanish-language versions of the assessments.

[15] Small FCCs were not included in the child assessment component of data collection because of the small number of children potentially available and likely to have parental consent.

Classroom Observations

To assess the quality of the ECE programs in the provider sample using alternative, validated tools not included in Delaware Stars, we observed classrooms in winter 2015 for all centers and FCCs in the provider sample. Observations were conducted for up to three classrooms per provider—one toddler classroom for two-year-olds, if such a room existed, and up to two preschool-age classrooms. For each classroom observation, we used three valid and reliable instruments, all of which have been used in prior QRIS validation studies: a global measure of ECE quality similar to the ERS used in the Delaware Stars rating scale, the Program Quality Assessment (PQA); and two focused measures of teacher-child interactions, the CLASS and CIS (see Table 2.4).

Together, these measures capture both structural aspects of program quality (e.g., dimensions of the learning environment and staff qualifications) and process components (e.g., dimensions of teacher-child interactions). Indeed, as shown in Table 2.5, the PQA components are aligned with the subscales in the ERS, as well as the four domains in Delaware Stars in which providers obtain points. The CLASS subscales and CIS are aligned with the Interaction subscale of the ERS, and, thus, they capture a narrow component of the quality dimensions incorporated in the Delaware Stars rating system. Nevertheless, given research showing a link between teacher-child interactions and child developmental outcomes, especially for the CLASS Instructional Support

Table 2.4. Measures of Classroom Quality

Measurement Tool	What It Measures
Preschool Program Quality Assessment (PQA)–Second Edition (High/Scope Educational Research Foundation, 2003, 2009; Hohmann, Lockhart, and Montie, 2013)	Global measure of quality collected by classroom observation and interview (teacher, director) • Center infant-toddler and preschool versions measure learning environment, daily routines, teacher-child interaction, curriculum planning and child observation/assessment, parent involvement and family services, staff qualifications and development, and program management • FCC version measures learning environment, daily routines, teacher-child interaction, and safety and health of the environment • 5-point scale
Classroom Assessment Scoring System (CLASS) (Pianta, La Paro, and Hamre, 2008; La Para, Hamre, and Pianta, 2012)	Measure of teacher-child interactions collected by observation for a classroom for use in center- or home-based settings • Toddler version has two subscales: Emotional and Behavioral Support and Engaged Support for Learning • Pre-K version has three subscales: Classroom Organization, Emotional Support, and Instructional Support • 7-point scale
Arnett Caregiver Interaction Scale (CIS) (Arnett, 1989)	Measure of teacher-child interactions collected by observation at the teacher level for use in center- or home-based settings • 26 items measure dimensions of positive relationships, caregiver punitiveness, caregiver permissiveness, and caregiver detachment • 4-point scale

Table 2.5. Comparison of Delaware Stars Rating Components and Alternative Measures of Quality

Delaware Stars Components	PQA Subscales	CLASS Subscales	CIS
ERS (e.g., ECERS)			
Space and Furnishings	✓ Learning Environment		
Language and Reasoning	✓ Learning Environment ✓ Adult-Child Interaction		
Activities	✓ Daily Routine		
Interaction	✓ Adult-Child Interaction	✓ Classroom Organization ✓ Emotional Support ✓ Instructional Support	✓ Positive Relationships ✓ Caregiver Punitiveness ✓ Caregiver Permissiveness ✓ Caregiver Detachment
Program Structure	✓ Daily Routine		
Domains			
Family and Community Partnerships	✓ Parent Involvement and Family Services		
Qualifications and Professional Development	✓ Staff Qualifications and Staff Development		
Management and Administration	✓ Program Management		
Learning Environment and Curriculum	✓ Learning Environment ✓ Curriculum Planning and Assessment		

subscale (Mashburn et al., 2008), it is important to determine whether the Delaware Stars ratings are reflecting variation in process quality as measured by aspects of teacher-child interactions.

These data for providers and enrolled children were supplemented with two additional sources of information:

- **Delaware Stars administrative data.** In addition to the information from the database of licensed providers, we also collected administrative data from OEL and the Delaware Department of Education for providers in Delaware Stars. Program ratings were captured for two points in time: August 2014, when providers were recruited into the study, and August 2015, the end point of our data collection. Because ratings could change during the course of our data collection, our preferred model specifications use the Stars rating as of August 2015, the measure of quality that is most likely to reflect program quality during the 2014–2015 year, when children were assessed. However, we also used the August 2014 rating as a sensitivity check. The Delaware Stars administrative data also included the subcomponents of the Stars rating: the ERS scores for observed classrooms and the number of points obtained in each of the four domains. The administrative data further indicated which providers had obtained their rating through an alternative certification pathway (as described in Chapter 1).

- **Census data.** In addition, we drew on publicly available 2009–2013 American Community Survey (ACS) data at the zip-code level to identify the characteristics of the neighborhoods in which the ECE providers in our sample were located and the characteristics of the neighborhoods where children in the sample resided. Appendix D lists the ACS indicators that we generated.

Characteristics of the ECE Providers in the Study

Descriptive statistics in Tables 2.6 and 2.7 provide a portrait of the ECE providers in the study according to information provided by the director or FCC provider. In both tables, we present unweighted percentages to illustrate the variation in the characteristics of the providers in our

Table 2.6. Characteristics of Programs in the Provider Sample

Characteristic	Unweighted Percentage
Program type[a] (percentage distribution)	
Small FCC	20.4
Large FCC	2.8
Center	70.2
School-based	6.6
Delaware Stars rating as of August 2015[a] (percentage distribution)	
Not in Delaware Stars	8.8
Starting with Stars	1.1
Star 2	17.7
Star 3	12.2
Star 4	34.3
Star 5	26.0
Program organizational form (percentage distribution)	
For profit, independent	48.0
For profit, chain	7.3
Not for profit	34.6
Public agency	10.1
Missing	*1.1*
Program enrolls children with POC (percentage distribution)	
POC not accepted	16.0
POC accepted	37.6
POC Plus accepted	31.5
Other, including both POC and POC Plus accepted	14.9
NAEYC accredited	20.4
Receives Head Start funding	9.9
Receives ECAP funding	13.3

SOURCE: Authors' analysis of provider sample director interviews and Delaware Stars administrative data.
NOTES: *N* = 181. Percentage distributions are calculated excluding missing cases. Percentage distributions may not sum to 100 because of rounding.
[a] Source is Delaware Stars administrative data.

Table 2.7. Characteristics of Children Enrolled in Programs: Director Interview Responses

Characteristic	Unweighted Percentage
Total program enrollment (percentage distribution)	
1–15 children	22.3
16–40 children	21.1
41–80 children	26.5
81 children or more	30.1
Missing	*8.3*
Enrollment by type (average percentage)	
POC	57.9
ECAP	5.7
Head Start	9.5
Racial-ethnic distribution of enrollment (average percentage)	
White, non-Hispanic	38.5
African American, non-Hispanic	40.8
Hispanic or Latino	13.3
Asian	1.7
Other	4.3
Percentage of enrolled children whose primary language is not English	
Average percentage	11.1
Percentage distribution	
0 percent	37.2
1 to 5 percent	27.8
6 to 10 percent	8.9
11 to 20 percent	11.1
21 to 100 percent	15.0
Missing	*0.6*

SOURCE: Authors' analysis of the provider sample director interviews.
NOTES: *N* = 181. Total program enrollment was collected via a separate program enrollment counts form. Percentage distributions are calculated excluding missing cases. Percentage distributions may not sum to 100 because of rounding.

sample.[16] For example, a little more than three-quarters of the providers in the study were centers or school-based programs, and the balance were small or large FCCs. Over half of the programs (60 percent) had a rating of Star 4 or 5 as of August 2015. Reflecting the transitory nature of Starting with Stars (which programs are expected to remain at for approximately three months), only two programs in the study had this rating by the end of the data collection period (August 2015). About 9 percent of the sample (16 programs) were not in Delaware Stars. Over half (55 percent) were for profit. A majority of the programs (84 percent) accepted children who received POC subsidies. Finally, as reported by providers, about 20 percent were NAEYC accredited, 10 percent received Head Start funding (either exclusively or along with other public subsidies, such as POC), and 13 percent received ECAP funding (again, either exclusively or in combination with other funding).[17]

[16] We do not present weighted results because the low response rate for providers not in Delaware Stars means that even weighted results do not fully represent the underlying population of all licensed providers in Delaware. Weighted results are most relevant when we consider providers in Delaware Stars, as we do for most results in Chapter 5, for example.

[17] In Chapters 3 and 4, we use Delaware Stars administrative data to determine which providers are rated through an alternative pathway either because they were a public school 619 program, a stand-alone Head Start or ECAP program, or an NAEYC-accredited program (see Table 1.2).

In terms of provider reports regarding the characteristics of the enrolled children, about one-fifth of the programs in the study enrolled 15 or fewer children, another one-fifth enrolled 16–40 children, a quarter enrolled 41–80 children, and the balance enrolled 81 or more children. The smallest program enrolled three children, and the largest enrolled 341 children. On average, the percentage of enrollees receiving POC subsidies was a slight majority. The average program had about equal shares of enrollees who were white and African American; the average share of Hispanic children was 13 percent.[18] The average percentage of students who spoke a language other than English at home was 11 percent, although about one-third of programs in the study reported enrolling no children who were non-English speakers. (Additional characteristics of the provider sample are reported in Appendix D.)

Characteristics of the Children Assessed for the Study

In large FCCs and centers, RAND assessed up to 15 children per ECE program across three kindergarten entry cohorts:

- **Youngest cohort** (born September 1, 2011, to August 31, 2012). These children were toddlers during the data collection period, ranging in age from two to three. They would be eligible to enter kindergarten in fall 2017.
- **Middle cohort** (born September 1, 2010, to August 31, 2011. These children were considered young preschoolers, at ages three and four during data collection. They would be eligible for kindergarten in fall 2016.
- **Oldest cohort** (born September 1, 2009, to August 31, 2010). These children were ages four and five during the data collection period and were eligible to enter kindergarten in fall 2015.

A total of 1,123 children were assessed in fall 2014 and again in spring 2015 with one or more developmental measures. Because we did not develop a frame of all enrolled children in the eligible age cohorts in the sampled providers, it was not possible to develop analytic weights to use in our analyses to reweight the sample of children to the statewide population of children in ECE programs. To gauge the representativeness of the children assessed, Table 2.8 shows the unweighted distribution of children assessed by age cohort, race-ethnicity, family income, and English-language status. We also show the distribution of children in Delaware ages 0 to 5, according to the ACS, for these same characteristics (except age cohort, because we did not have birthdates). (Additional child characteristics are reported in Appendix D.)

[18] Note that Table 2.7 is reporting on the average, across programs, in the distribution of child characteristics. Even with weighting, this is not the same thing as the distribution of those characteristics across all children enrolled in programs in Delaware.

Table 2.8. Characteristics of Children Assessed with One or More Developmental Measures in Fall 2014 and Spring 2015

Characteristic	Unweighted Percentage in Child Sample	Characteristics of Delaware Children Younger Than Age 6 in ACS
Age cohort		
Youngest cohort (enter kindergarten in fall 2017)	21.6	–
Middle cohort (enter kindergarten in fall 2016)	38.2	–
Oldest cohort (enter kindergarten in fall 2015)	40.2	–
Race-ethnicity		
White, non-Hispanic	45.2	49.6
African American, non-Hispanic	24.1	23.9
Hispanic or Latino	16.6	14.5
Asian, non-Hispanic	3.1	4.3
Other, non-Hispanic	8.8	7.6
Missing	2.1	–
Family income		
$25,000 or less	29.2	19.9
$25,001–$50,000	15.3	20.1
$50,001–$100,000	18.3	31.7
$100,001 or above	22.4	28.4
Missing	*14.8*	–
Child's primary language is not English	15.2	26.8
Number of children	1,123	68,226

SOURCE: Authors' analysis of child sample parent survey and 2014 ACS One-Year Public Use Microdata Sample (PUMS).
NOTES: Delaware characteristics are from the 2014 ACS One-Year PUMS. Regarding the primary language spoken at home, the ACS measures whether the household language for children under 6 years of age is not English. – = not available or not applicable.

As shown in Table 2.8, 22 percent of our child sample fell in the youngest kindergarten entry cohort, and the remainder of the sample was almost equally divided between the two older cohorts. Consistent with the statewide Delaware child population (whether in formal care or not), almost half of assessed children were white, one-quarter African American, and 15 percent Hispanic. One-third of assessed children came from families who earned less than $25,000, and only one-quarter came from families who earned $50,000–$100,000. Relative to the statewide population, the child sample has a proportionately larger share of low-income children, which could be attributable to the population of children in formal care at these ages, as well as possible selectivity of children for whom we received parental consent and were able to complete two waves of assessments. We also note that family income was not reported by the parent for 15 percent of our child sample.

For children assessed in both fall and spring, 6.2 months elapsed between the two assessments, on average. Table 2.9 reports the average gain from fall to spring in the six child assessment measures. In each case, the average gain score is significantly different from zero ($p < 0.0001$), indicating that children, on average, experienced positive developmental gains over the course of the school year when the assessments took place. To interpret the magnitude of the

Table 2.9. Mean Fall-to-Spring Change Scores on Child Assessment Measures

Full Sample	*N*	Mean	SD	t-statistic	*p*	Effect Size
PPVT	988	8.04	10.73	23.54	< 0.0001	0.75
WJ–Letter Word Identification	1,076	13.26	18.28	23.78	< 0.0001	0.73
WJ–Applied Problems	998	9.28	18.67	15.69	< 0.0001	0.50
HTKS	789	7.36	14.08	14.69	< 0.0001	0.52
DECA–Absence of Behavior Problems	819	20.15	11.24	51.32	< 0.0001	1.79
DECA–Total Protective Factors	969	10.18	10.41	30.43	< 0.0001	0.98

SOURCE: Authors' analysis of child sample assessments and Delaware Stars administrative data.
NOTES: A total of 1,123 children had at least one developmental assessment measured in fall and spring. The effect size is the ratio of the mean to the standard deviation (SD), also known as Cohen's *d*.

gains across the assessments when using different scales, Table 2.9 also reports the effect size (the ratio of the mean to the standard deviation). The largest gains occur in the DECA measures of social-emotional and behavioral skills (effect sizes of one standard deviation or higher). Gains in the language and literacy measures (PPVT, WJ–Letter Word Identification) are about three-quarters of a standard deviation, while those for early mathematics skills and executive function are about one-half of a standard deviation.

Chapter 3. Delaware Stars Ratings and Program Quality

As discussed in Chapter 1, one key objective of our evaluation is to examine the relationship between Delaware Stars ratings and ECE program quality. In addition, we are interested in how program characteristics, beyond those captured in the Delaware Stars ratings, relate to program quality. In particular, we aim to address the following two questions (from Table 1.1):

> *Question 1*. To what extent do the quality tiers of Delaware Stars accurately reflect differential levels of program quality such that programs at the top levels provide a higher-quality care and early learning experience than programs at lower levels?
>
> *Question 2*. What is the relationship between program characteristics and quality in Delaware Stars?

As part of the first question, because the QRIS rating is a summary measure that aggregates over an array of program characteristics (i.e., the ERS score and the specific program features assessed in the four quality domains in Table 1.3), we also seek to understand the relationship between components of the QRIS rating and program quality. Addressing the first question, in particular, has been a central part of most other state QRIS validation studies.

To address these questions, in this chapter we use data from our provider sample to compare Delaware Stars ratings (or the component parts of the ratings), as well as other program characteristics not included in the ratings, with program quality as measured by PQA, CLASS, and CIS. We collected these measures for 172 providers, although not all measures were completed for each program.[19] Before presenting our findings, we provide a brief overview of our methods, which are similar to those followed in other QRIS validation studies (e.g., those summarized in Table 1.6).

Methods

In addressing Question 1, our primary interest is to see how a measure of ECE program quality, such as the PQA, varies as programs move up the Delaware Stars rating scale. As illustrated in Table 3.1, one approach is to compare the average PQA score (or the score on each of the CLASS subscales or the CIS) for programs in our provider sample based on their Delaware Stars rating status. And because the PQA and CLASS are classroom-based measures of quality, we first calculate the provider-level average PQA or CLASS score across all observed classrooms. The CIS is first averaged across providers in the same classroom, and then the provider-level average across all classrooms is obtained.

[19] One or both of the CLASS measures (CLASS Pre-K or CLASS Toddler) were collected for all 172 programs. We have valid PQA scores for 162 providers and CIS scores for 167 providers.

Table 3.1. Program Quality Score Summary Statistics by Delaware Stars Rating Status

Provider Quality Measure and Delaware Stars Status	Number of Programs	Summary Statistics			
		Mean	Standard Deviation	Minimum	Maximum
PQA					
Not in Delaware Stars	13	3.22	0.67	1.75	4.33
Starting with Stars/Star 2	31	2.98	0.52	1.96	3.96
Star 3	19	3.18	0.54	2.31	3.95
Star 4	57	3.39	0.51	2.37	4.64
Star 5	42	3.55	0.38	2.66	4.33
CLASS Pre-K Emotional Support					
Not in Delaware Stars	13	5.17	1.16	3.08	6.54
Starting with Stars/Star 2	32	4.97	0.95	3.00	6.68
Star 3	20	5.24	0.96	3.42	6.67
Star 4	60	5.61	0.71	3.58	7.00
Star 5	44	5.51	0.76	2.46	6.58
CLASS Pre-K Classroom Organization					
Not in Delaware Stars	13	4.58	1.46	2.22	6.67
Starting with Stars/Star 2	32	4.03	1.06	1.94	5.78
Star 3	20	4.29	1.14	1.83	6.06
Star 4	60	4.82	1.10	2.00	6.89
Star 5	44	4.76	1.03	1.63	6.48
CLASS Pre-K Instructional Support					
Not in Delaware Stars	13	2.07	0.44	1.22	2.78
Starting with Stars/Star 2	32	2.03	0.82	1.17	4.72
Star 3	20	1.83	0.43	1.00	2.56
Star 4	60	2.20	0.70	1.22	4.56
Star 5	44	2.59	0.90	1.17	5.22
CLASS Toddler Emotional and Behavioral Support	11	4.99	1.19	2.57	6.40
Not in Delaware Stars	28	4.88	1.03	2.83	6.33
Starting with Stars/Star 2	19	4.93	1.10	3.20	6.27
Star 3	34	5.50	0.98	3.43	7.00
Star 4	27	5.46	0.74	3.50	6.77
Star 5	11	4.99	1.19	2.57	6.40
CLASS Toddler Engaged Support for Learning					
Not in Delaware Stars	11	2.58	0.84	1.33	3.80
Starting with Stars/Star 2	28	2.35	0.81	1.22	4.40
Star 3	19	2.37	0.99	1.00	4.28
Star 4	34	2.85	1.13	1.00	5.44
Star 5	27	3.21	0.78	2.00	4.80
CIS					
Not in Delaware Stars	12	3.23	0.57	2.24	3.87
Starting with Stars/Star 2	32	3.08	0.45	1.98	3.73
Star 3	20	3.25	0.41	2.42	3.85
Star 4	59	3.35	0.35	2.38	3.92
Star 5	44	3.38	0.25	2.63	3.76

SOURCE: Authors' analysis of provider sample classroom observations and Delaware Stars administrative data.
NOTES: Results are weighted to account for provider sampling and nonresponse. Classroom quality scores are first averaged across observed classrooms for each program.

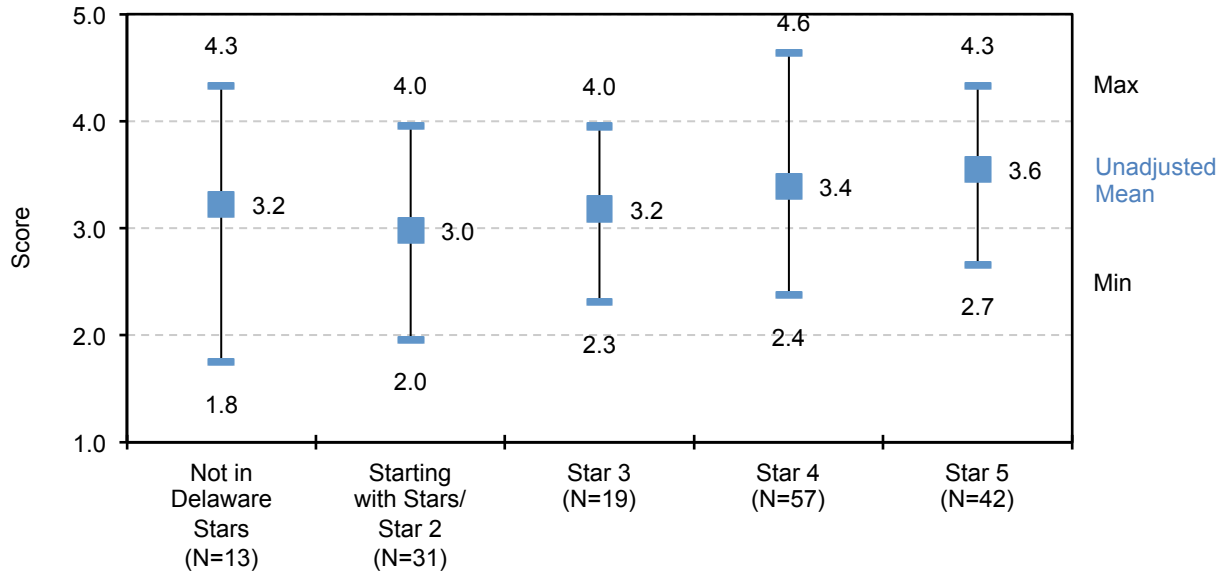
Note that in Table 3.1 we have grouped together Starting with Stars and Star 2 providers for two reasons. First, we have relatively few providers at these rating tiers, so we improve our sample-based estimate of the mean for those providers by increasing the number in the combined group. Second, providers at Starting with Stars and Star 2 have not undergone the full rating

process that occurs at Star 3 and above. In other words, those providers, because they are "in process," have not been rated based on classroom observations or verified on the points-based standards, as is done with providers at Star 3 and above (with the exception of alternative pathway providers). Thus, programs at Starting with Stars and Star 2 (and providers not participating in Delaware Stars at all) could include a mix of lower- and higher-quality providers.

For this reason, we would expect to see the strongest relationship between rating tiers and ECE program quality as the programs move from Star 3 to Star 4 and Star 4 to Star 5, as those programs have all been assessed using the ERS and points-based standards (with the exception of programs rated through an alternative pathway). According to Delaware Stars, these programs differ in their quality with each step up the rating scale by at least one scale point on the seven-point ERS scale (equivalent to a 0.7-point increase on the PQA five-point scale or a 0.5-point increase in the four-point CIS scale) and by a 20-point differential in the number and weight associated with the other program features captured in the points-based quality standards.[20]

Figure 3.1 shows the results in Table 3.1 specifically for the PQA. Like Table 3.1, these data are unadjusted for other program characteristics that explain some of the variation in PQA

Figure 3.1. PQA Scores for ECE Programs in the Provider Sample by Delaware Stars Rating



SOURCE: Authors' analysis of provider sample classroom observations and Delaware Stars administrative data.
NOTES: Results are weighted to account for provider sampling and nonresponse. Classroom quality scores are first averaged across observed classrooms for each program.

[20] For example, Delaware Stars programs with a Star 3 rating must have achieved an ERS score of at least 3.4, whereas those with a Star 4 or Star 5 rating had to achieve scores of at least 4.4 and 5.4, respectively. Although it may be possible to find a Star 3 program with an ERS score of 4.3 and a Star 4 program with an ERS of 4.4—so they are separated by just 0.1 scale points—as long as ERS scores are distributed along the interval that applies to each rating tier (i.e., 3.4 to 4.3, 4.4 to 5.3, and 5.3 to 7), we would expect programs, on average, at each rating tier to be different by about one scale point. The same logic applies to the scores in the four quality domains in which points are earned.

scores. Figure 3.1 provides a visual representation of the regression-based findings (where we do make those statistical adjustments) that we report in the rest of this chapter. Namely, Figure 3.1 shows that there is considerable overlap in centers' PQA scores regardless of their Star level. In other words, at each quality rating level, there is considerable variation in the PQA, as reflected in the minimum and maximum values. Interestingly, the minimum and maximum scores tend to increase when moving from one rating level to the next, but not always. Even so, providers with a Star 5 Delaware Stars rating may have scored from 2.7 to 4.3 on the 5-point PQA scale, which is not that much higher than the range of 2.3 to 4.0 for Star 3 providers. Excluding those programs that were not in Delaware Stars, there was a small increase of 0.1 to 0.3 points in the mean PQA score in each progressively higher rating level, from Starting with Stars/Star 2 to Star 5. Similar patterns hold for the CLASS and CIS measures of program quality included in Table 3.1.

The data shown in Table 3.1 and Figure 3.1 are unadjusted for other program characteristics that explain some variation in the outcomes (in that case, the PQA). To best isolate the relationship between the Delaware Stars rating and the quality measures we collected for this evaluation, we adopt a regression-based approach, equivalent to the descriptive analysis in Table 3.1, that produces adjusted mean quality scores (PQA, CLASS, or CIS) for each Delaware Stars rating tier. The regression-based approach allows us to control for other factors that might explain some of the variation in the quality measure beyond the Delaware Stars rating. In all cases, we aggregate PQA[21] and CLASS scores[22] across classrooms within centers. Because the CIS is a teacher-level measure, we first average CIS scores across teachers in the same classroom and then compute a center-level CIS score as the average of the classroom-level averages.

In results presented later in the chapter, we conduct statistical tests to determine whether the average quality score at each rating level is significantly different from the other rating levels. For example, is the mean PQA score at Star 5 significantly different from the mean PQA score at Star 4 or Star 3? A power analysis shows that our sample sizes provide power at 80 percent or higher to detect a large effect size of 0.8 in the mean quality scores. For the PQA, this translates into a scale-point difference of approximately 0.5. For the CLASS subscales other than the Instructional Support scale, we have power at 80 percent or higher to detect differences in average quality across rating tiers of about one scale point. For Instructional Support, we can

[21] In the case of the PQA, although there are separate versions for center- and home-based providers and for classrooms serving toddlers versus preschool-age children, we pool all 181 providers together in the analysis and treat the PQA as one common measure.

[22] In the case of the CLASS, a single aggregate score is typically not reported, but rather the subscales are the measures of interest. However, because the number of subscales and constructs for the subscales differ for the Toddler and Pre-K versions of the CLASS, it is not possible to pool results. Thus, results for the CLASS Toddler measure are analyzed for those providers with one or more toddler or mixed-age groups at their site. The CLASS Pre-K measure is likewise reported for the subset of providers with one or more preschool- or mixed-age classrooms.

detect differences of about 0.4 scale points. For the CIS, we have power of 80 percent or higher to detect a difference of 0.5 points. In each case, these are substantively meaningful differences in program quality. For example, as noted earlier, the Delaware Stars rating tiers from Star 3 to Star 4 and Star 4 to Star 5 are based on a scale-point increase in the ERS, which is the same absolute change on the seven-point CLASS scale, a 0.7 scale-point increase in the PQA, and a 0.5 scale-point increase in the CIS.

All results have been weighted to generalize to the population of licensed Delaware ECE programs. Appendix E provides more detail on the models and control variables, with detailed results presented in Appendix F. We also report on alternative specifications in Appendix G. We generally found that our preferred specifications were robust. For example, instead of aggregating across classrooms in the center using the average, we also examined relationships based on the lowest-scoring classroom and also using classrooms as the unit of analysis. In both cases, we did not find different relationships from those we report in this chapter. In addition, we estimated models excluding programs rated through an alternative pathway (because they do not undergo the full rating process), and the findings were unchanged.

Because we were testing many hypotheses, we also adjusted our tests for statistical significance to reduce the chance that we found statistically significant differences by chance, given the large number of tests we performed. Appendix E describes this adjustment in more detail. We highlight in the narrative only those results that remained statistically significant after correcting for multiple hypotheses testing.

In addition to the summary Delaware Stars rating, we also examined whether there was a relationship between components of the rating scale (e.g., the ERS score, the points in each of the four quality domains, whether an essential standard is met, or the sum of the points associated with all essential standards that were met) and the PQA, CLASS, or CIS. In doing so, we must account for the changes over time in the rating structure associated with these rating components. As explained in Appendix E, the general approach was to first stratify programs in the provider sample and standardize their points-based measures before pooling the programs together to examine their subscores in relation to our alternative quality measures.[23] We did this because the amount of points differs by FCCs versus centers, and because programs were verified under two different sets of Delaware Stars program rules. After standardizing the subscores, we then examined whether programs scoring in the lowest to highest quartile of a given component of the Delaware Stars rating (e.g., the Management and Administration domain) obtained higher average scores on the PQA, CLASS, or CIS.

To address Question 2, we adopt a similar approach but instead focus on the relationship between ECE program characteristics not included in the Delaware Stars ratings and alternative

[23] For example, centers that were last validated under the old set of standards obtained 8–13 points within the Management and Administration domain if they were in the lowest quartile of programs; 14–16 points in that domain if they were in the second-lowest quartile of programs; 17–18 points in that domain if they were in the second-highest quartile; and 19–20 points in that domain if they were in the highest quartile.

measures of program quality. As part of this analysis, we consider provider type, subsidy density, and alternative pathway status, among other characteristics. For this analysis, we required that we had at least ten providers with a given characteristic. This means that we excluded large FCCs from consideration, as well as several types of providers rated through an alternative pathway.

Relationship Between Delaware Stars Ratings and Alternative Measures of Program Quality (Question 1)

We focus first on the relationship between the overall Delaware Stars ratings and alternative measures of program quality before turning to a similar analysis based on the components of the Delaware Stars ratings. Because the PQA captures all of the dimensions of quality incorporated into the Delaware Stars rating (see Table 2.5), we expected average PQA ratings to have the strongest relationship with Delaware Stars ratings. The CLASS subscales and CIS capture aspects of teacher-child interactions that are measured, in part, through the ERS. Nevertheless, it is important to understand whether increases in Delaware Stars ratings are associated with increases in the quality of teacher-child interactions.

Alternative Quality Measures Modestly Rise as Delaware Stars Ratings Increase

As shown in Table 3.2, as Stars ratings rise from Star 3 to Star 5, so did most of the regression-adjusted average scores on seven different quality measures—the PQA, the three CLASS Pre-K

Table 3.2. Regression-Adjusted Average Score on Alternative Quality Measures by Delaware Stars Rating Level

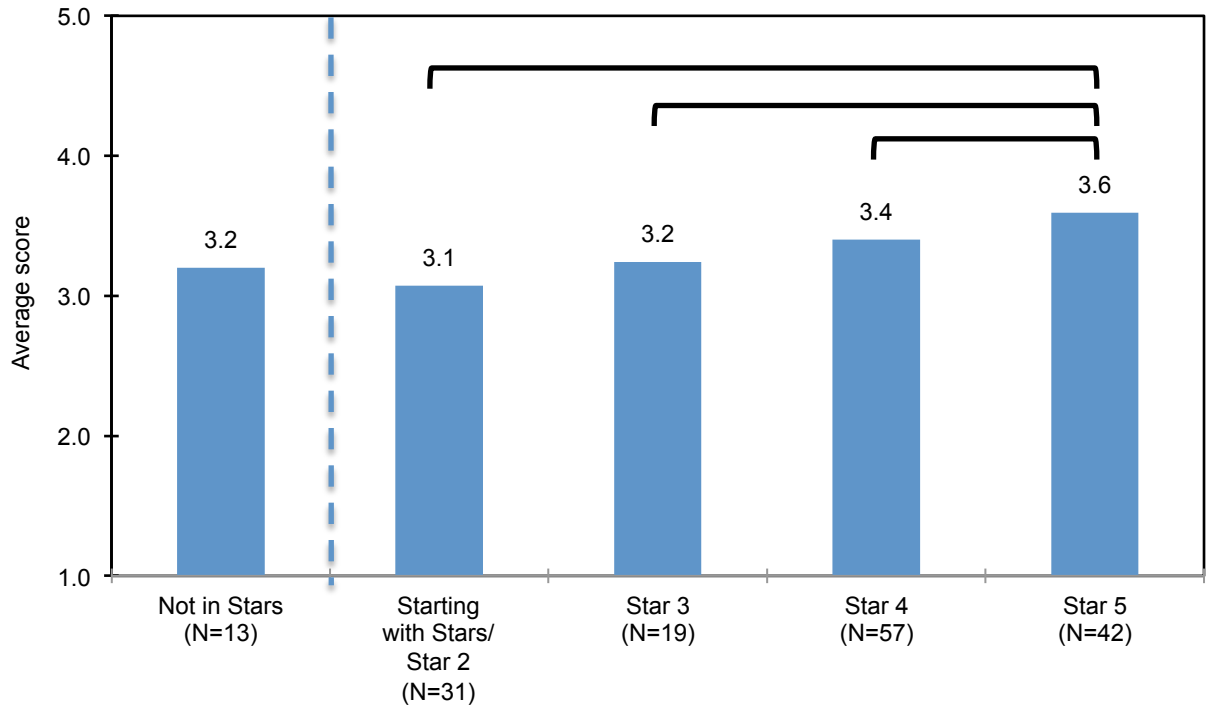
Alternative Program Quality Measure	Not in Stars	Starting with Stars or Star 2	Star 3	Star 4	Star 5
PQA	3.2[e]	3.1[E]	3.2[E]	3.4[E]	3.6[a,B,C,D]
CLASS Pre-K Emotional Support	5.1	5.6	5.4	5.7	5.6
CLASS Pre-K Classroom Organization	4.4	4.6	4.6	5.0	4.8
CLASS Pre-K Instructional Support	2.0[E]	2.3[C]	1.8[B,E]	2.0[E]	2.5[A,C,D]
CLASS Toddler Emotional and Behavioral Support	5.0	4.8	5.0	5.4	5.6
CLASS Toddler Engaged Support for Learning	2.5[e]	2.5[e]	2.6[e]	2.8	3.2[a,b,c]
CIS	3.2	3.3	3.3	3.3	3.4
Number of programs					
Where PQA observations completed	13	31	19	57	42
Where CLASS Pre-K observations completed	13	32	20	60	44
Where CLASS Toddler observations completed	11	28	19	34	27
Where CIS observations completed	12	32	20	59	44

SOURCE: Authors' analysis of provider sample classroom observations and Delaware Stars administrative data.
NOTES: Superscript letters denote that the difference in the regression-adjusted average score is significantly different at $p < 0.05$ prior to adjusting for multiple hypothesis testing from [a] not in Stars, [b] Starting with Stars or Star 2, [c] Star 3, [d] Star 4, and [e] Star 5. Capitalized superscript letters denote differences that remained statistically significant after adjusting for multiple hypothesis testing.

subscales, the two CLASS Toddler subscales, and the CIS. In most cases, providers not in Stars and at Starting with Stars or Star 2 show similar results on average, with mean scores that are also similar to Star 3 programs. However, the increase in average classroom scores on the PQA, CLASS, or CIS from one rating level to the next is generally modest and most often not statistically significant. Further, the alternative classroom quality measures did not always increase in lockstep with the increase in the rating level. Figures 3.2 and 3.3 present two out of the seven sets of results contained in Table 3.2. We selected those results where the positive relationship was strongest between the Stars rating levels and the alternative classroom quality measures.

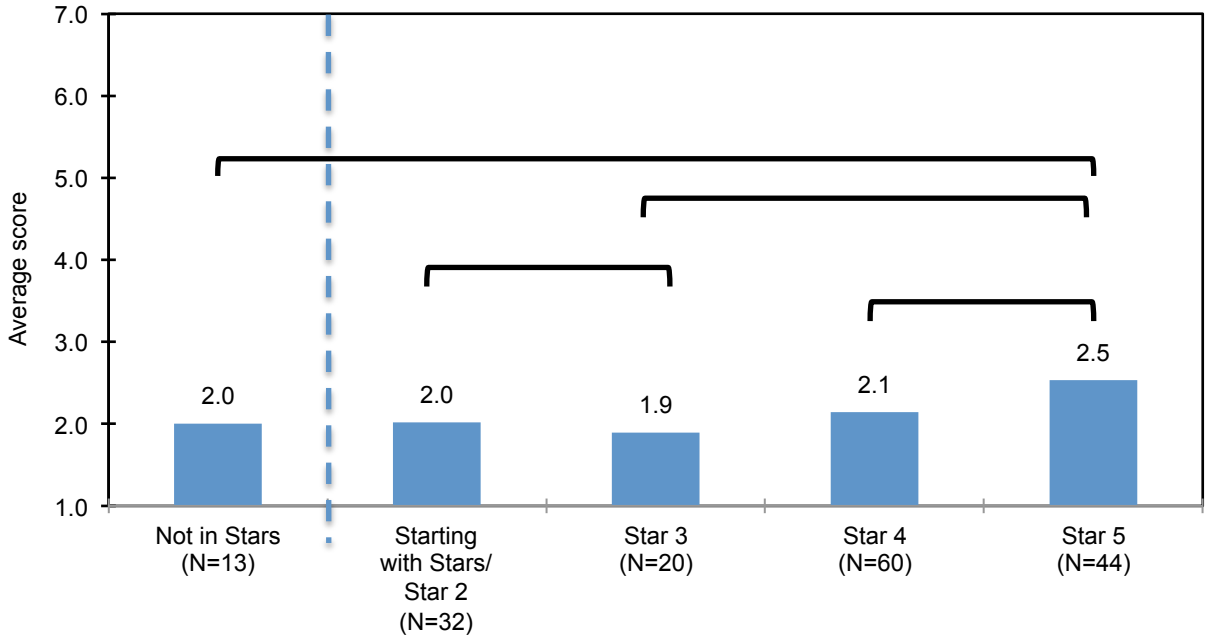
As anticipated, we observe the strongest relationship between our classroom observation scores and Star 3 to Star 5. This is because only programs at Star 3 and above (that are not certified using an alternative pathway) have surpassed the required minimum ERS classroom score that would theoretically be correlated with the PQA and potentially with the CLASS and CIS measures as well. As noted earlier, programs not in Delaware Stars and programs at Starting with Stars and Star 2 did not obtain ERS classroom observations.

Figure 3.2. Relationship Between Delaware Stars Rating and PQA Score



SOURCE: Authors' analysis of provider sample classroom observations and Delaware Stars administrative data.
NOTES: Results shown in Table 3.2. Average scores by Star level are regression-adjusted from models described in Appendix E. Brackets show pairwise comparisons that are significantly different from one another at $p < 0.05$ after adjusting for multiple hypothesis testing. The number of providers at each rating level is shown in parentheses.

Figure 3.3. Relationship Between Delaware Stars Rating and CLASS Pre-K Instructional Support Score



SOURCE: Authors' analysis of provider sample classroom observations and Delaware Stars administrative data.
NOTES: Results shown in Table 3.2. Average scores by Star level are regression-adjusted from models described in Appendix E. Brackets show pairwise comparisons that are significantly different from one another at $p < 0.05$ after adjusting for multiple hypothesis testing. The number of providers at each rating level is shown in parentheses.

While we often observe the expected positive stair-step pattern between Delaware Stars ratings and increased average classroom quality scores, the size of the increase in average scores of the alternative measures is smaller than the equivalent of the one-point increase that the Delaware Stars system requires for the ERS to move from Star 3 to Star 4 and from Star 4 to Star 5. For example, average PQA classroom observation scores increased by 0.2 points with each Star-level increase, far less than the 0.7-point equivalent to a scale-point increase in the ERS. Indeed, the difference in the average PQA scores between the Star 3 and Star 5 programs on the PQA is 0.4 points on the five-point scale, the equivalent of a 0.6 scale-point change in the ERS.

In the leftmost bar of Figures 3.2 and 3.3 are the average classroom quality scores for the 13 programs in our study that are not in Delaware Stars. We hypothesize that programs not in Delaware Stars are likely a combination of some of the lowest- and highest-quality ECE providers in the state—those programs that feel they are not likely to receive a high rating and those programs that feel they do not need the quality rating to attract parents. Similarly, programs at Starting with Stars and Star 2 are also likely a mixture of higher- and lower-quality providers because they are programs that have newly entered Delaware Stars. Some programs may readily attain a Star 3 or higher rating, while other programs may stay at a Star 2 rating for some time.

Finally, the results in Table 3.1 demonstrate that the level of quality attained by Star 5 programs in Delaware is often below the level that would be considered high quality. This is particularly evident for the CLASS Pre-K Instructional Support subscale, which reaches 2.5 on the seven-point scale for Star 5 programs. As noted in Chapter 1, effective preschool programs, such as those in Tulsa and Boston, achieve averages of 3.2 to 4.3 on the Instructional Support measure. Thus, there is scope for improving the level of quality at the highest rating tier, which would further help to differentiate it from the next-lowest rating level.

These findings comport with most of the prior QRIS validation studies summarized in Table 1.6, which likewise also found a positive, but weak, relationship between alternative quality measures and higher QRIS rating tiers. The relatively low level of quality even at the highest rating tier is another finding in common with other validation studies. Finally, we note that our findings were not affected by the exclusion of programs rated through an alternative pathway (see Appendix G and Table G.1).

Alternative Quality Measures Are Not Positively Related to Scores on the Components of the Overall Delaware Stars Rating

To assess whether any one part of the Delaware Stars ratings bore a relationship with our seven quality scores (PQA, CLASS, and CIS), we examine the four domain scores, as well as their total ERS score, for programs with Star 3 to Star 5 ratings.[24, 25] As shown in Figures 3.4 and 3.5, as well as Table F.1, we almost always observed no consistent pattern in average quality scores as the number of points that a program obtained in a given domain of Delaware Stars rose. In some cases, average quality scores were higher in the top score ranges than in bottom score ranges. For example, in the Qualifications and Professional Development domain, the CLASS Pre-K Instructional Support scores and CLASS Toddler Engaged Support for Learning scores generally rose as the domain points increased. Likewise, in the Learning Environment and Curriculum domain, the PQA and CLASS Pre-K Instructional Support scores generally rose as the domain points increased.

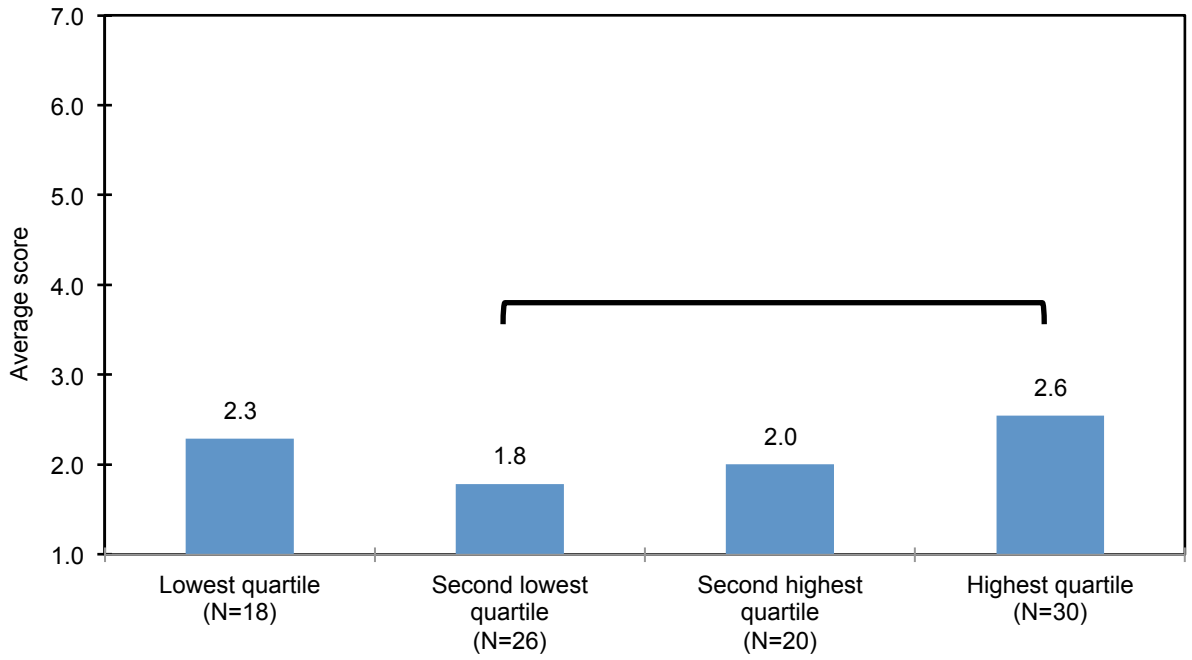
---

[24] As described in Chapter 1, ECE programs must both obtain a score on the ERS classroom observations that exceeds a minimum threshold (for their lowest-scoring classroom) *and* obtain additional points within the following four domains: (1) Family and Community Partnerships, (2) Qualifications and Professional Development, (3) Management and Administration, and (4) Learning Environment and Curriculum. The higher the star level, the higher the ERS score threshold and the greater the number of points required in each of the four domains.

[25] For this analysis, we first ranked each program's relevant domain score or ERS composite score within the overall distribution of validated program scores and categorized programs into quartiles of the distribution—i.e., either in the lowest fourth, second-lowest fourth, second-highest fourth, or top fourth of program points for that category. Since the total number of available points differed across FCCs versus centers and differed by the old versus revised Delaware Stars standards, we separately ranked programs' validation points and ERS composite scores within each of those four categories to which the program's most recent validation belonged. Note that fewer programs are included in these analyses of Delaware Stars domain scores than are shown in the results presented in Figures 3.2 and 3.3 because, by definition, only programs at Star 3 and above possess validation and ERS data. In addition, per program rules, some types of alternative certification pathway programs lack ERS and/or validation data.

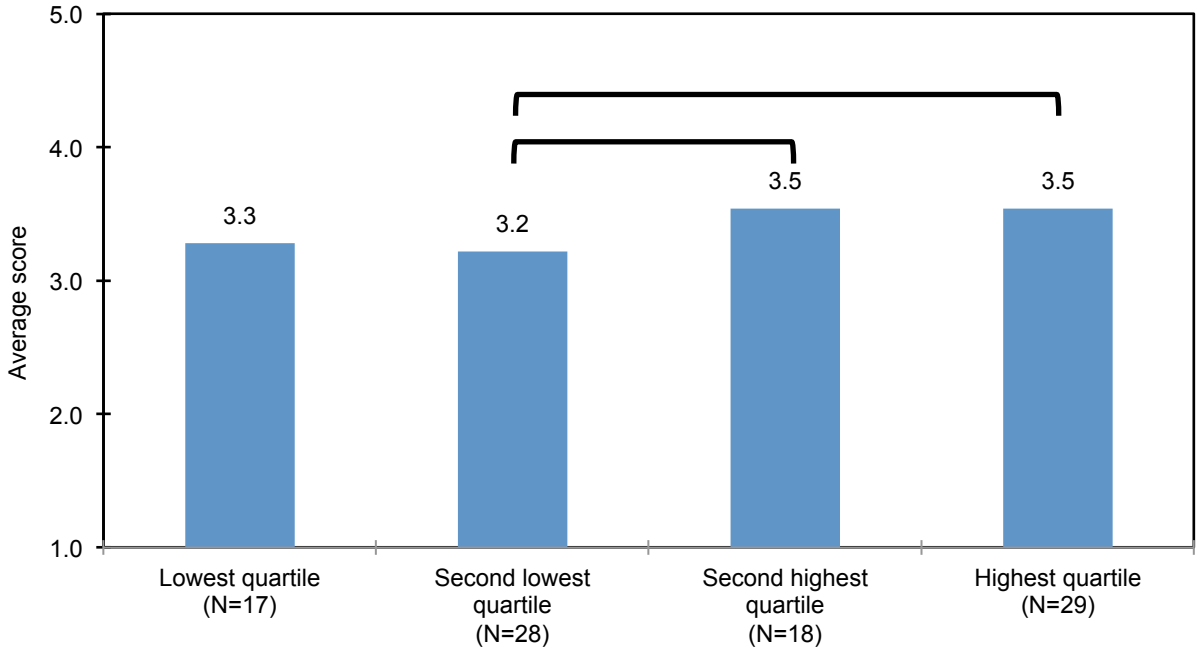
In Figures 3.4 and 3.5, we present two instances where the domain scores were significantly and mostly positively related to a subset of the alternative quality measures. For example, in Figure 3.4, we see a very modest U-shaped curve. It counterintuitively shows that those programs scoring lowest on the Management and Administration domain obtained a higher average Instructional Support score on the CLASS Pre-K instrument than on the next two higher-scoring tiers within the Management and Administration domain. We do, however, observe the expected positive stair-step in the second lowest quartile through the highest-scoring programs (highest quartile) on the Management and Administration domain. As shown in Figure 3.5 for the PQA score, we see a modest but positive relationship according to successively higher Learning Environment and Curriculum domain scores.

Figure 3.4. Relationship Between Management and Administration Domain Score and the CLASS Pre-K Instructional Support Score



SOURCE: Authors' analysis of provider sample classroom observations and Delaware Stars administrative data.
NOTES: Results shown in Table F.1. Sample restricted to programs at Star 3 to Star 5, excluding those rated through an alternative pathway. Average scores by quartile of Management and Administration points-based score are regression-adjusted from models described in Appendix E. Brackets show pairwise comparisons that are significantly different from one another at $p < 0.05$ after adjusting for multiple hypothesis testing. The number of providers at each rating level is shown in parentheses.

Figure 3.5. Relationship Between Learning Environment and Curriculum Domain Score and PQA Score



SOURCE: Authors' analysis of provider sample classroom observations and Delaware Stars administrative data.
NOTES: Results shown in Table F.1. Sample restricted to programs at Star 3 to Star 5, excluding those rated through an alternative pathway. Average scores by quartile of Learning Environment and Curriculum points-based score are regression-adjusted from models described in Appendix E. Brackets show pairwise comparisons that are significantly different from one another at $p < 0.05$ after adjusting for multiple hypothesis testing. The number of providers at each rating level is shown in parentheses.

As shown in Tables F.2 and F.3, there was no statistically significant relationship between any one of the six essential standards (or the sum of the points associated with the essential standards) and alternative quality measures after adjusting for multiple hypothesis testing.

Relationship Between Program Characteristics and Alternative Measures of Program Quality (Question 2)

As part of this analysis, we examined the relationship between the PQA, CLASS, and CIS and measures of program characteristics not included in the Delaware Stars ratings, such as program license type, Stars Plus cohort status, whether rated through an alternative pathway, and subsidy density. As noted earlier, because of sample size limitations, we could not consider all characteristics. (Table F.4 provides a descriptive comparison of mean scores on the independent quality measures and various program characteristics.)

Some Program Characteristics are Related to Alternative Measures of Quality

We found small, statistically significant differences in alternative quality measures across different types of programs participating in Delaware Stars (see Tables F.5 and F.6). Namely, after controlling for other program characteristics, we found that small FCC programs scored higher on average on the PQA and on the Emotional Support CLASS subscale scores for the Toddler instrument. However, because invited FCCs participated at a lower rate in our evaluation than did centers, we believe that the higher FCC ratings might reflect sample selectivity. In other words, higher-performing FCCs might have been the ones most likely to consent to participate in the evaluation.

While Head Start stand-alone programs, rated through the alternative pathway, also scored higher on the PQA than regular-pathway Delaware Stars programs, the differences were not statistically significant after adjusting for multiple hypothesis testing. With one exception, programs serving a majority of low-income children either through POC or ECAP tended to score lower on alternative quality ratings, but the differences were not statistically significant after adjusting for multiple hypotheses testing. The exception was programs where a majority of children received ECAP, which scored lower on the CLASS Instructional Support measure compared with regular-pathway Delaware Stars programs. Stars Plus programs also scored lower on the PQ and on one of the CLASS subscales, but the differences were not significant after adjusting for multiple hypothesis testing.

Conclusion

Consistent with other QRIS validation studies, we found that average quality scores on alternative measures generally increased as Delaware Stars rating levels increased, although at a small rate and generally not with statistically significant increases from one star level to the next. The exceptions were statistically significant but small increases from Star 3 to Star 5 in average PQA scores—the most comprehensive quality measure that captured the key quality constructs rated in Delaware Stars—and in average CLASS Pre-K Instructional Support scores—a key measure of teacher-child interactions. At the same time, the level of quality as measured by PQA and the CLASS, even for Star 5 programs, was below the level expected for high-quality programs, and the incremental improvements in quality in moving from Star 3 to Star 5 were small according to these measures.

When we examined the relationship between our alternative measures of quality and the components that constitute the Delaware Stars ratings, we found inconsistent results. The expected relationships were strongest for the Qualifications and Professional Development domain, as well as the Learning Environment and Curriculum domain, but only for some of the alternative quality measures. Finally, we found that FCCs slightly outscored centers on alternative quality ratings, while programs with a majority of ECAP enrollees scored lower than regular-pathway programs on one of the seven quality measures collected.

With 181 programs in the validation study, the sample was not sufficiently large to detect small differences in program quality across rating tiers with statistical precision. Nevertheless, the number of participating ECE programs was sufficiently sized to detect substantively meaningful differences in quality, such as a scale-point or larger change on a seven-point scale (or a 0.5 scale-point change on a five-point scale), as would be expected in a rating system like Delaware Stars, where the minimum ERS quality rating score incorporated within Stars must increase by a point on the seven-point ERS scale to rise from one Star level to the next.

Chapter 4. Delaware Stars Ratings and Child Developmental Outcomes

In this chapter, we examine whether provider Delaware Stars ratings and the components of the rating are associated with the developmental outcomes of the children enrolled in their program. Specifically, we examine whether a program's Delaware Stars rating as of August 2015 is positively associated with children's performance as of spring 2015 on a series of early academic, cognitive, social-emotional, and behavioral assessments, taking into account prior performance on the same assessments as of fall 2014, as well as other child and program characteristics. We answer the following research questions:

> *Question 3*. All else being equal, do young children participating in higher-rated programs have better learning and developmental outcomes than similar children in nonparticipating or lower-rated programs?
>
> *Question 4*. What dimensions of Delaware Stars program ratings are most vital to child learning and developmental outcomes?

As part of Question 3, we also consider whether children in low-income families who participate in higher-rated programs benefit more than children in low-income families in lower-rated programs. Question 3 has been the central focus of a number of other QRIS validation studies, including those listed in Table 1.7. Question 4 goes beyond considering the Stars rating as a whole to examine whether any of the components that are used to derive the overall Stars rating are associated with children's developmental outcomes.

We address these questions by combining data from the provider sample with the child sample, which allows us to examine children's learning over the 2014–2015 academic year as a function of a program's Delaware Stars rating or the components of the rating. As part of our analyses, we also examine the relationship between children's development and the alternative measures of program quality (i.e., the PQA, CLASS, and CIS) we collected for the provider sample. Because we did not assess children in small FCCs and just 2 percent of the child sample were enrolled in large FCCs, our findings in this chapter are most applicable to center-based programs. Before presenting our findings, we briefly describe our analytic approach, which is similar to the approach taken in other recent QRIS validation studies that include child outcomes (e.g., the studies listed in Table 1.7).

Methods

To answer these questions, we specified a preferred regression model to examine the relationship between Delaware Stars ratings (or the ratings components) with children's assessed scores on early academic, cognitive, social-emotional, and behavioral measures in spring 2015. Our

preferred specification is set out in Appendix E, as are alternative models to check the robustness of the results from the main model. As described in Appendix G, our substantive findings were not altered with any of these different empirical methods. More-detailed findings for this chapter are documented in Appendix F. As we did in Chapter 3, we describe only results that remain statistically significant after correcting for multiple hypothesis testing.

As noted in Chapter 1 in the context of similar validation studies, when examining the relationship between child outcomes and Delaware Stars ratings, we need to account for the selection of children into ECE programs. For example, if more advantaged families select into higher-performing ECE programs, a positive relationship between program quality and child development may be explained, in part, by the contribution of family resources to children's learning. Following the methods used in other QRIS validation studies, such as those in Table 1.7, our models of child developmental outcomes in spring 2015 on any given assessment include controls for the child's prior performance (using the fall 2014 assessment), the observed child and family characteristics based on information provided by parents and the provider, the neighborhood characteristics based on the child's home zip code, and other observed characteristics of the ECE program. Using our regression-based approach, we present in this chapter adjusted mean child assessment scores for each Delaware Stars rating tier, controlling for other factors that explain some of the variation in children's spring 2015 developmental outcomes. Although we screened children for English-language ability and then assessed children in Spanish on parallel assessments, we excluded Spanish-language results from the analyses. However, we retained English-language results from children who were also assessed in Spanish, along with an indicator for primary language at home. (As a sensitivity analysis, we estimated models with only those children who spoke English at home.)

To help the reader compare results across child developmental outcomes that are measured on different scales, we present standardized results.[26] Thus, the reader can interpret the difference in outcomes between Delaware Stars rating levels as an effect size, which is a metric that can be compared across research studies and across the six developmental assessments analyzed in this chapter. These comparisons can be made across columns in the table or across the bar heights in the figures. To present results on a uniform absolute scale, the figures throughout this chapter have a y-axis range of one standard deviation unit.

Relationship Between Delaware Stars Overall Ratings and Children's Development (Question 3)

In Table 4.1, we present the relationship between the three assessments of children's early academic skills, the measure of executive function, and the measures of children's social-

[26] We did this by first predicting average children's outcomes by Delaware Stars rating level using regression estimates and then dividing those predicted scores by the standard deviation of the scores.

emotional and behavioral skills relative to the Delaware Stars rating level of the ECE program in which the child was enrolled.

Differences in Children's Development Across Rating Tiers Are Generally Small and Mostly Statistically Insignificant

As shown in the first three rows of Table 4.1, we did not find a statistically significant, positive relationship between Delaware Stars rating tiers and children's early literacy or early mathematics skills. The lack of a relationship is also evident for the measure of executive function (HTKS) and the two DECA measures of social-emotional and behavioral skills. Again, Star 3 and above are the levels at which we expect to see the strongest relationship with child learning because it is only at those levels that programs are validated on the ERS and other Delaware Stars standards.

Figure 4.1 illustrates the results for the WJ–Letter Word Identification assessment. The differences in the bar heights range from 0.02 standard deviations (i.e., 10.71 for Star 4 versus 10.73 for Star 5) to 0.12 standard deviations (programs not in Stars or Star 3 versus Star 5). There is the expected stair-step pattern in moving from Star 3 to Star 4 and Star 5, but the difference in performance from Star 3 to Star 5 is just 0.12 standard deviations. Further, the bar height is greatest for programs that are in Starting with Stars and Star 2, those that have yet to be formally assessed with the ERS and quality standards. The relatively small sample size at the Star 3 rating level means that we cannot say with confidence that children's development is higher at Star 4 and Star 5 programs compared with Star 3 programs.

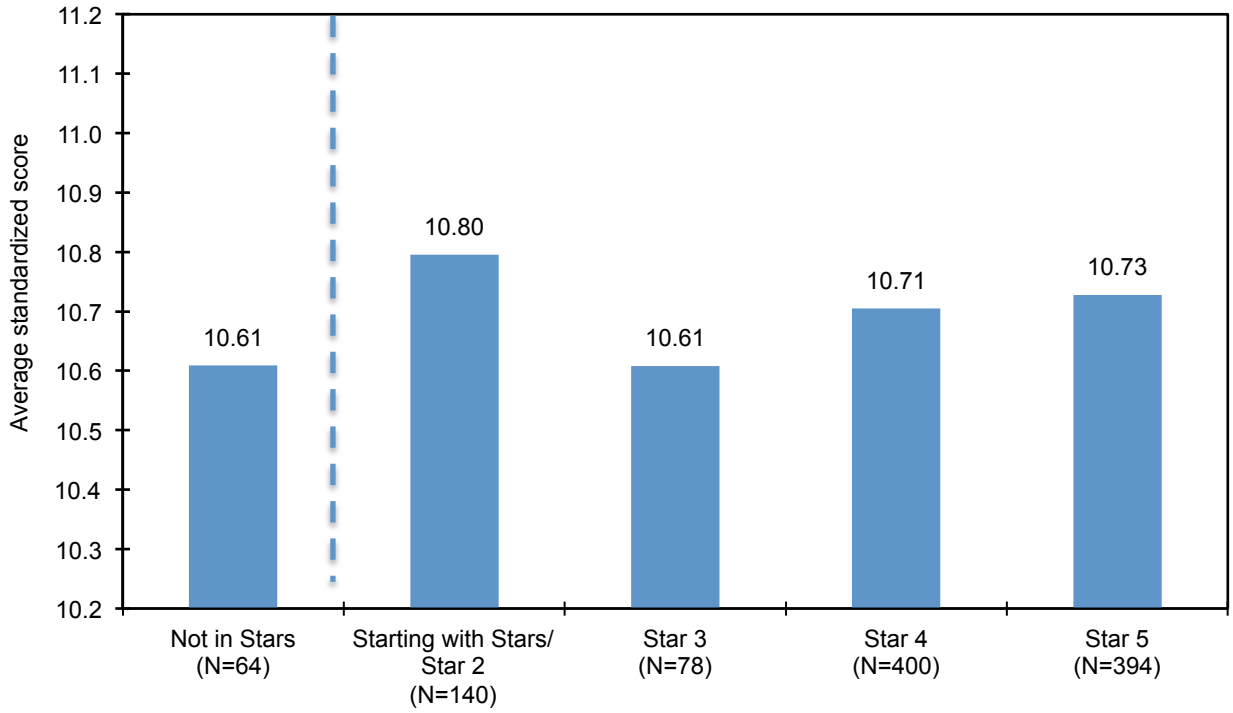
In the case of the HTKS—a measure of executive function that assesses attention, inhibitory control, and working memory—children in Star 5 programs had significantly higher performance than those in Starting with Stars and Star 2 programs, a difference of 0.34 standard deviations (see Figure 4.2). But there were no significant differences after adjusting for multiple hypothesis

Table 4.1. Regression-Adjusted Average Child Developmental Assessment Standardized Score by Delaware Stars Rating Level

Developmental Outcome (Spring 2015)	Not in Stars	Starting with Stars or Star 2	Star 3	Star 4	Star 5
PPVT	6.41	6.42	6.38	6.35	6.37
WJ–Letter Word Identification	10.61	10.80	10.61	10.71	10.73
WJ–Applied Problems	15.22	15.28	15.01	15.11	15.15
HTKS	0.85	0.69[E]	0.92	0.85[e]	1.03[B,d]
DECA–Absence of Behavior Problems	5.56	5.27	5.22	5.36	5.48
DECA–Total Protective Factors	6.32	5.96	5.89	5.77	5.91

SOURCE: Authors' analysis of child sample assessments and Delaware Stars administrative data.
NOTES: See Table F.7 for the number of children assessed. Superscript letters denote that the difference in the regression-adjusted average score is significantly different at $p < 0.05$ prior to adjusting for multiple hypothesis testing from [a] not in Stars, [b] Starting with Stars or Star 2, [c] Star 3, [d] Star 4, and [e] Star 5. Capitalized superscript letters denote differences that remain statistically significant after adjusting for multiple hypothesis testing.

Figure 4.1. Relationship Between Delaware Stars Rating and Child Standardized WJ–Letter Word Identification Score



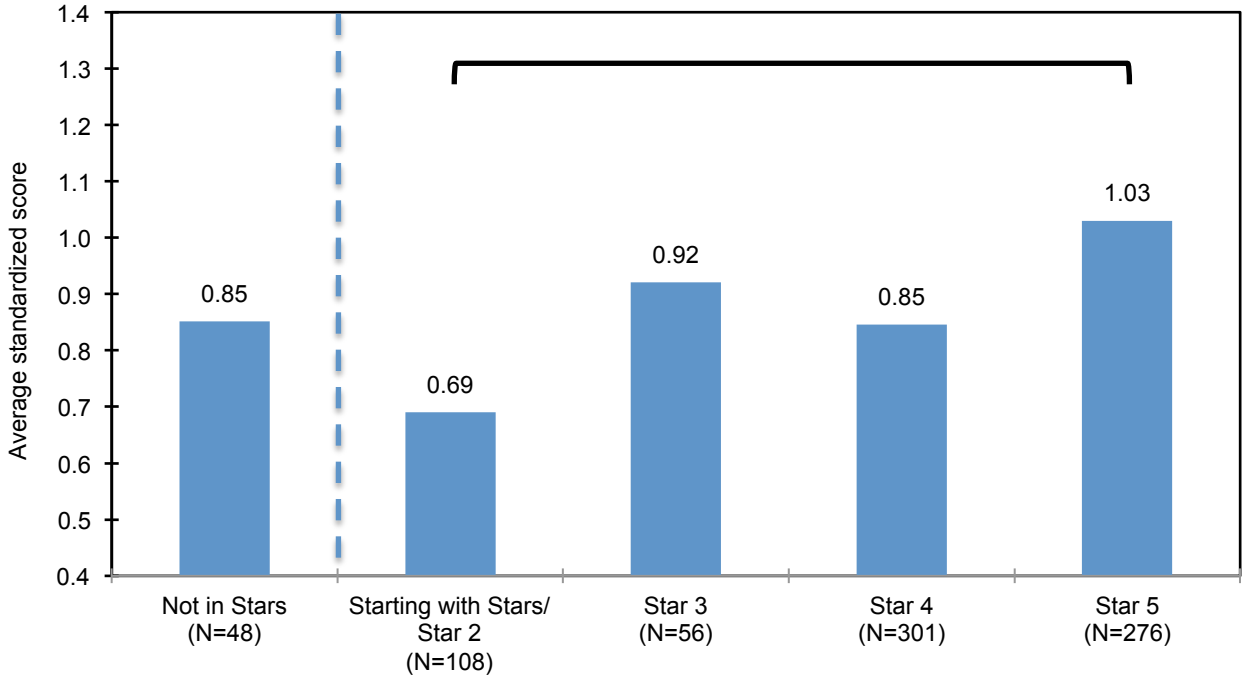
SOURCE: Authors' analysis of child sample assessments and Delaware Stars administrative data.
NOTES: Results shown in Table 4.1. Average assessment scores by star level are regression-adjusted from models described in Appendix E. Brackets show pairwise comparisons that are significantly different from one another at $p < 0.05$ after adjusting for multiple hypothesis testing. The number of children assessed at each rating level is shown in parentheses. WJ–Letter Word Identification w-scores are divided by the standard deviation. The overall mean Letter Word Identification standardized w-score is 10.71.

testing between Star 3, Star 4, and Star 5 programs. In fact, the bar height for Star 3 exceeds that for Star 4, so the expected stair-step pattern across programs with formal ratings is not evident.

The DECA measure of an absence of behavioral concerns shows the expected stair-step pattern (see Figure 4.3), with a difference in performance of 0.26 standard deviations between Star 3 and Star 5 that is again not statistically significant. The highest bar in this case is for children not in Delaware Stars.

In sum, our estimates show average performance increasing when moving from Star 3 to Star 4 to Star 5 on two of the three early academic assessments—WJ–Letter Word Identification (Figure 4.1) and WJ–Applied Problems (Table 4.1)—and on the DECA assessment indicating an absence of behavior problems (Figure 4.3). The differences from Star 3 to Star 5 are 0.12, 0.14, and 0.26 standard deviations, respectively. The difference in performance by rating levels is statistically significant only in the case of the HTKS, and then only between Starting with Stars and Star 2 programs and Star 5 programs (Figure 4.2).

Figure 4.2. Relationship Between Delaware Stars Rating and Child Standardized HTKS Score



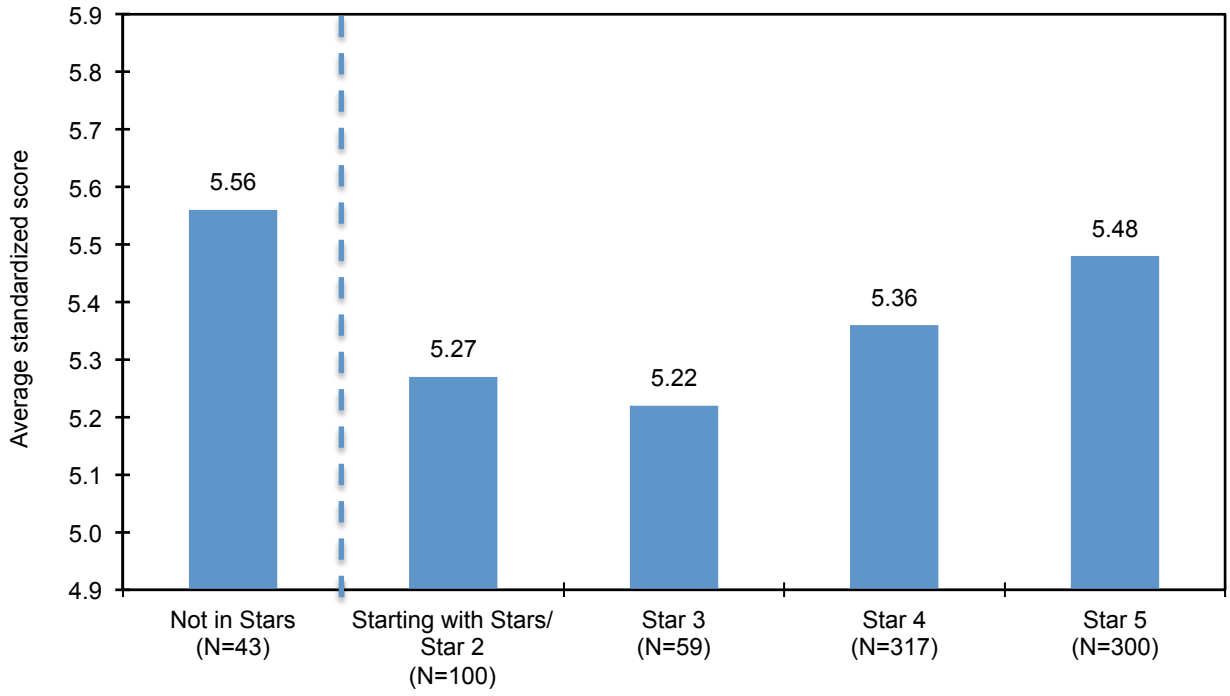
SOURCE: Authors' analysis of child sample assessments and Delaware Stars administrative data.
NOTES: Results shown in Table 4.1. Average assessment scores by star level are regression-adjusted from models described in Appendix E. Brackets show pairwise comparisons that are significantly different from one another at $p < 0.05$ after adjusting for multiple hypothesis testing. The number of children assessed at each rating level is shown in parentheses. HTKS raw scores are divided by the standard deviation. The overall mean HTKS raw standardized score is 0.89.

Children from Low-Income Families Do Not Experience Differential Learning in Higher-Rated Programs When Compared with Lower-Rated Programs

We also examined whether low-income children in particular benefited from enrolling in high-rated versus low-rated programs.[27] We generally found no statistically significant difference in the scores of low-income children in low- versus high-rated programs. There was one suggestive exception: Teachers reported greater protective factors for low-income children in programs rated in the top two rating tiers (Star 4 or Star 5) than the bottom three rating tiers (Starting with Stars to Star 3), but the difference was not statistically significant after adjusting for multiple hypothesis testing.

[27] We defined children from low-income families as those who met any one or more of the following four criteria: Their family reported an income of $25,000 or less; the director reported that the child received ECAP; the director reported that the child received the POC subsidy; or the child was enrolled in a Head Start center. Half (50 percent) of the children we assessed qualified as being from low-income families.

Figure 4.3. Relationship Between Delaware Stars Rating and Child Standardized DECA Absence of Behavioral Concerns Score



SOURCE: Authors' analysis of child sample assessments and Delaware Stars administrative data.
NOTES: Results shown in Table 4.1. Average assessment scores by star level are regression-adjusted from models described in Appendix E. Brackets show pairwise comparisons that are significantly different from one another at $p < 0.05$ after adjusting for multiple hypothesis testing. The number of children assessed at each rating level is shown in parentheses. DECA–Absence of Behavioral Concerns t-scores are divided by the standard deviation. The overall mean Absence of Behavioral Concerns standardized t-score is 5.39.

Relationship Between Delaware Stars Ratings Components and Children's Development (Question 4)

Two Domains of Delaware Stars Are Predictive of Modest Positive Differentials in Selected Children's Outcomes

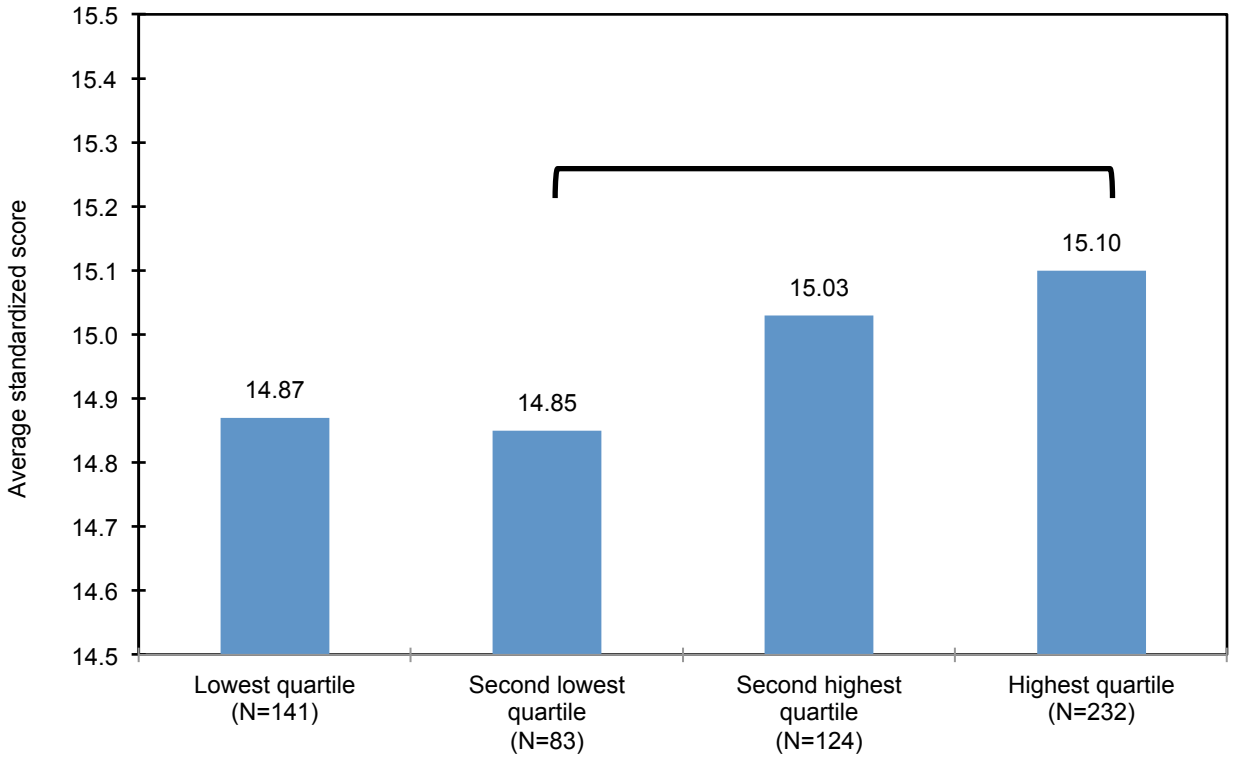
The relationship between children's developmental outcomes and the four domain scores and the ERS score that contribute to the overall Delaware Stars rating was even more attenuated than the relationship found between these domain scores and alternative quality measures. Specifically, we found no statistically significant relationship between children's developmental outcomes and scores on either the Family and Community Partnerships domain or the Learning Environment and Curriculum domain (see Table F.8). We also found no relationship between children's outcomes and a program's ERS score. We did find, however, that scoring high on the Qualifications and Professional Development domain was related to modestly increased average WJ–Letter Word Identification scores (not graphed) and Applied Problems scores (see Figure

4.4), with differences in performance between lower-scoring and higher-scoring programs equal to 0.18 and 0.25 standard deviations, respectively. We also found a nonlinear but statistically significant relationship between Management and Administration scores and children's performance on HTKS (Figure 4.5), with a contrast in performance of 0.32 standard deviations between programs scoring in the lowest quartile and the highest quartile of the domain.

The Bundle of Essential Standards Is Associated with Children's Executive Function

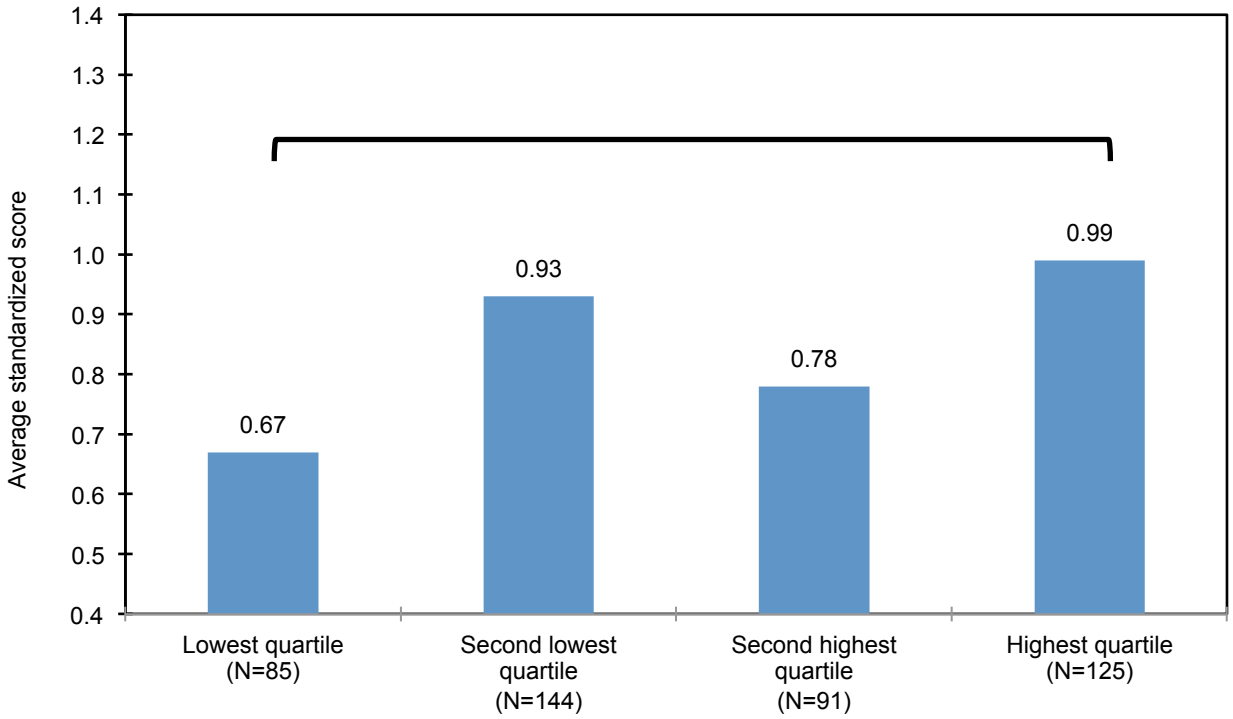
We found that the sum of the points that programs obtained on each of the six essential standards was predictive of moderate differences of approximately 0.33 of a standard deviation on the measure of executive function (see Table F.10). Separately, none of the six essential standards was associated with children's outcomes after accounting for multiple hypothesis testing (see Table F.9).

Figure 4.4. Relationship Between Qualifications and Professional Development Domain Score and Child Standardized WJ–Applied Problems Score



SOURCE: Authors' analysis of child sample assessments and Delaware Stars administrative data.
NOTES: Results shown in Table F.8. Sample restricted to programs at Star 3 to Star 5, excluding those rated through an alternative pathway. Average scores by quartile of Qualifications and Professional Development points-based score are regression-adjusted from models described in Appendix E. Brackets show pairwise comparisons that are significantly different from one another at $p < 0.05$ after adjusting for multiple hypothesis testing. The number of children assessed at each rating level is shown in parentheses. WJ–Applied Problems w-scores are divided by the standard deviation. The overall mean Applied Problems standardized w-score is 15.00.

Figure 4.5. Relationship Between Management and Administration Domain Score and Child Standardized HTKS Score



SOURCE: Authors' analysis of child sample assessments and Delaware Stars administrative data.
NOTES: Results shown in Table F.8. Sample restricted to programs at Star 3 to Star 5, excluding those rated through an alternative pathway. Average scores by quartile of Management and Administration points-based score are regression-adjusted from models described in Appendix E. Brackets show pairwise comparisons that are significantly different from one another at $p < 0.05$ after adjusting for multiple hypothesis testing. The number of children assessed at each rating level is shown in parentheses. HTKS raw scores are divided by the standard deviation. The overall mean HTKS standardized score is 0.87.

Conclusion

The findings of our analysis of the relationship between developmental outcomes for children primarily in center-based programs and Delaware Stars ratings provide a mixed picture. On the one hand, for the children in our sample, we found some developmental measures—both early academic skills and social-emotional and behavior skills—where children's performance was higher in Star 4 and Star 5 programs than in Star 3 programs. But we had a small group of children in Star 3 programs, so the statistical tests indicate that it is unclear whether these differences accurately reflect true differences or whether the specific sample of children in our study in Star 3 programs happened to have lower scores compared with the larger groups of children in Star 4 and Star 5 programs. It is possible that we would find a different result if we had a different sample or a larger sample of children in Star 3 programs. The HTKS measure of executive function was the only one to show a significant difference across rating tiers and then only between Starting with Stars and Star 2 programs that had yet to be formally rated and Star 5

programs with the highest rating. This is potentially a meaningful finding in that executive function skills are predictors of school readiness, academic achievement, and long-term health and educational outcomes (McClelland et al., 2014).

Considering the components that make up the Delaware Stars ratings, we found that points awarded in two of four quality domains and the sum of points that programs obtained on the bundle of essential standards were predictive of performance on some of the developmental assessments, in some cases early academic skills and in other cases executive function. There was no evidence that top-rated Delaware Stars programs confer greater benefits for low-income children as compared with lower-rated programs.

These mixed findings are consistent with those reported for the validation studies of other statewide QRISs, such as those in Colorado, Minnesota, and Wisconsin, reported in Table 1.7. These results from the validation portion of our overall evaluation also comport with the results from the virtual pilot study that we described in our Year 1 report, in which we estimated the effects of a QRIS with Delaware Stars' general design on a national sample of children's outcomes. We note that, like the suggestive findings for Delaware Stars, some QRIS validation studies have found small or modest associations of higher QRIS ratings with children's early academic skills, which forms the basis for a recommendation that we include in Chapter 6 to consult other QRIS results to further refine Delaware Stars.

Chapter 5. Delaware Stars System Performance

In this chapter, we turn to the remaining five study questions, listed in Table 1.1. These questions pertain to quality improvement in Delaware Stars, financial incentives, and other aspects of system performance. Specifically, we seek to address two questions related to quality improvement:

> *Question 5*. To what extent does the TA (i.e., on-site support, orientation, etc.) provided to Delaware Stars participants help providers to move up in Delaware Stars?
>
> *Question 6*. To what extent do high-need programs that participate in Stars Plus enhance their program quality?

Three other questions concern financial incentives, consumer perspectives, and systems operations:

> *Question 7*. Are the financial incentives and supports for providers sufficient to support the needed quality improvements?
>
> *Question 8*. What do consumers understand about Delaware Stars? Do consumers ask about ratings?
>
> *Question 9*. How well do the Delaware Stars system components operate?

In the Year 1 and Year 2 reports of our evaluation (Schwartz et al., 2014; Auger, Karoly, and Schwartz, 2015), we addressed various aspects of these questions, drawing on a round of interviews in the first year of the evaluation with Delaware Stars system administrators, along with provider and parent focus groups. Administrative data from the QRIS provided the basis for analyses of participation in Delaware Stars, movements across rating tiers over time, receipt of financial incentives, and receipt of TA support. We summarize the findings from these earlier analyses where relevant. In addition, we draw on quantitative and qualitative findings from the interviews of program directors we conducted during the course of our 2014–2015 data collection. The perceptions and experiences of program leaders—both those participating in Delaware Stars and those not currently participating in the QRIS—provide valuable insights about the QRIS as a whole.

Methods

The methods for any findings based on the first two evaluation reports are detailed in those documents, so we do not repeat them here. As described in Chapter 2, the director interviews consisted of structured interviews with directors of programs in the provider sample, first in fall

2014 and again in spring 2015.[28] Using a scripted questionnaire with both close-ended and open-ended responses (included in Appendix C), we interviewed directors in 181 licensed small FCCs, large FCCs, centers, and school-based preschools, of which 165 were in Delaware Stars and 16 were not in Delaware Stars (see Table 2.5). We conducted these interviews primarily by phone, although some were conducted in person. The fall portion of the interview focused on program characteristics, and the spring portion covered perceptions about Delaware Stars, the value of Delaware Stars, and financial incentives. Where we report tabulations of interview responses in this chapter, we apply the provider sample weights to account for sampling and nonresponse. In this way, responses are representative of all licensed providers in Delaware. Most questions were relevant only for providers participating in Delaware Stars, so tabulations of those questions exclude the nonparticipating providers and will be representative of all providers participating in Delaware Stars.

Technical Assistance Within Delaware Stars (Question 5)

As we documented in the Year 2 report, TA is the second-largest financial investment Delaware makes in its QRIS. (The state's largest financial investment in Delaware Stars is the POC Tiered Reimbursement Bonuses, whereby Star 3 to Star 5 providers receive additional payments for children with POC subsidies based on their rating level.) Our analysis of the Delaware Stars TA records from October 2013 through September 2014—which is the year leading up to the 2014–2015 period when we collected data from the provider sample and child sample—indicated that ECE programs received the intended number of TA visits as part of Delaware Stars. On average, providers at Star 2 to Star 5 (that were not special cases, such as the Stars Plus cohort or those rated through an alternative pathway) received about 14 to 24 onsite visits annually. These visits amounted to 18 to 39 hours of TA supports in a year. As intended, visits and annual hours were lower for Starting with Stars programs, and they were the highest for Stars Plus programs that received weekly TA. For regular Delaware Stars programs, the number of visits and annual hours peaked at the Star 3 level and then tapered off at the Star 5 level.

To glean more information about those TA visits, we asked the directors of programs in Delaware Stars about their perceptions of TA. Directors of 151 programs in Delaware Stars responded to our questions about TA. Specifically, in addition to asking about receipt of various types of specialty TA (described later on), we posed the following three open-ended questions to directors in Delaware Stars about TA:

- Does your Delaware Stars technical assistance provider offer guidance and feedback that you view as improving your program?
- How do you think Delaware Stars technical assistance could be improved, if at all?

[28] At 13 out of 181 programs, we interviewed curriculum coordinators or other similarly senior administrators. For convenience, we refer to all of the interviewees as "directors" throughout this chapter.

- For those who indicated receiving specialty technical assistance: Do you think that specialized technical assistance has helped to improve your program? Which types of specialty TA, and why or why not?

Directors Had Positive Views of TA Overall, with Some Qualifications

Nearly all (90 percent) of the directors of the 152 programs who responded to TA questions answered positively. For those who elaborated beyond simple responses of "yes," directors tended to comment on TA providers' quick responses, TA providers offering concrete guidance to improve classroom practices, and TA providers helping with paperwork or more generally helping to navigate the Delaware Stars system. Examples of typical positive comments include the following: "Every time she comes she gives me a pointer and gives me time to correct it. She always talks to me about it and helps me improve." "The TA [provider] constantly provides verbal and written feedback by letting us know how we are doing, where we need to go and what we need to purchase."

Those who indicated that TA did *not* help to improve their program tended to focus on one or more of three themes:

- lack of knowledge on the TA provider's part about Delaware Stars (eight mentions)
- lack of responsiveness or lack of availability for prompt visits (five mentions)
- for programs that had more than one TA provider over time, contradictory or inconsistent guidance from the TA providers (four mentions).

These three themes appeared again in directors' suggestions for how to improve TA. Directors of 68 programs offered suggestions for how TA could improve. (Directors of another 80 programs answered that they either did not know how to improve TA or that there was no need for improvement of TA.) Three dominant themes emerged for improvement:

- a desire for **greater consistency** between state licensing, TA providers, validators, and ERS assessors about Delaware Stars rules, including greater knowledge of each program's rules (32 mentions)
- a desire for **greater frequency** of visits or more time from TA providers (29 mentions)
- a desire for **less turnover** among TA providers assigned to programs (seven mentions).

These views are illustrated in Text Box 5.1.

Text Box 5.1. Illustrative Responses from Directors Regarding Improvements to TA

Directors voice their desire for greater **consistency** among TA providers:

- "I guess it is just being clear and consistent in how the standards are being met. Based on [situations] when [the] TA [provider] approves, but the validators say you haven't been approved."
- "Consistency and validity for the pre-assessments so that we are getting the correct information."
- "The TA [provider] needs to be more knowledgeable about the upcoming changes to DE Stars. The DE Stars program is constantly changing and the [TA providers] need to be informed about it."
- "She's nice but not that knowledgeable. Always willing to help but often didn't know the answers to questions."
- "Could provide more continuity of guidance when there is a change in TA staff."
- "Everyone makes mistakes and leaves things out. Being more thorough and having a checklist. It was daunting. Maybe if she had a checklist when she first comes out."
- "Because I have multiple centers, they should be the same TA [provider]. They're so different."

Directors speak about **greater frequency** of TA provider visits:

- "It would be helpful if the [TA providers] could come visit the site more often and help schedule time that works for the teachers instead of being on a rotation."
- "We would like to have more time with them if that was possible."
- "If they could give a more set schedule for follow through. Stay full day in the centers."
- "More regular visits. The TA [provider] only comes 1 time a month. I would prefer 3 times a month."
- "When we need to reschedule, it takes around 45–60 days to reschedule a visit."
- "If we didn't have to wait so long for an appointment, then it would be nice if they were more available."

Directors speak about TA provider **turnover**:

- "In two years, we have had 4 [TA providers] and one substitute. Every time you get a new TA [provider] you have to figure out where you are in their calendar."
- "They help you out when they have cohorts or time, but then they tend to disappear. Like you get attached to one and then they have to go—not a lot of continuity, so you can't build a relationship with them."
- "I think within the last 5 years we've had almost 7 [TA providers]. It's always changing and some are great and some aren't great."

SOURCE: Director interviews, spring 2015.

Directors Were Extremely Positive About the Value of Specialty TA and Specialized Training

Starting in 2014, the state introduced specialized supports for quality improvement, either through specialty TA or though specialized training for ECE program staff. The specialty TA focused on four topics: infant and toddler care, the ERS, health and nutrition, and early childhood mental health. In addition, specialized training was made available for ECE program staff about the use of the Ages and Stages developmental screening tool and child developmental assessment. Generalist TA providers make recommendations to directors about particular specialty TA and specialized training that their program would benefit from, or programs may request the specialized support themselves. Almost all directors who had reported obtaining TA through Delaware Stars also reported receiving at least one type of specialized TA or training (see Table 5.1). By far, most common was receipt of training on the Ages and Stages developmental assessment tool (63 percent of programs). One in five to one in three programs participated in the other types of specialty TA. In addition, nearly half of program directors reported participating in the Aim4Excellence leadership course offered through Delaware Stars.

Referring to their experiences with these specialized quality improvement services, the directors strongly endorsed their value overall. Almost all directors (107 out of 116) who responded to our question said that the specialized TA or training helped to improve their program. Most directors went on to comment on one or more of the six areas of specialized support. Three of the six forms of specialized support received the highest praise.

- The **early childhood mental health consultancy** was most frequently mentioned (31 mentions). Many directors used words like "fabulous" and reported that the consultant specifically improved the dynamics of classrooms or behaviors of individual children. For example, one director said, "Mental health one was so awesome and helped a couple

Table 5.1. Responses by Directors of Programs in Delaware Stars Regarding Specialized Assistance

Survey Item	Weighted Percentage of Respondents
Ever receipt of specialized TA or training (more than one may apply)	
Infant/toddler care specialist	32.6
ERS specialist	34.5
Health and nutrition specialist	18.8
Early childhood mental health consultation	25.2
Training for Ages and Stages developmental screening	62.8
Training for child assessment	31.1
Director participated in Aim4Excellence course	48.7

SOURCE: Authors' analysis of provider sample director interview responses.
NOTES: Based on 156 respondents. Results are weighted to account for provider sampling and nonresponse.

kids here and my whole preschool program." Only four out of 31 mentions were negative—e.g., "too little, too late, too infrequent," as explained by one director.

- **Ages and Stages** training was also a frequently named support (23 mentions), and all but one director said that the training was helpful. Many said that it was the most helpful of the six types of support. For example, one director noted, "I was able to use it on a particular kid who was two, who was below his developmental level. I was able to refer to a guardian who referred to a doctor and now he has outpatient therapy and a specialized IEP." Several directors called out its value in helping programs improve communication with and involvement of parents.
- **Infant/toddler care specialty TA** was equally frequently named (mentioned 23 times), and all but three directors viewed the training positively. Most comments related to the theme that the TA provider offered hands-on help that yielded improved class environments. A typical comment here included the following: "Was having a rough time with my infant/toddlers, and having a TA [provider] come by made a big difference in the infant toddler room." Or "improving our interaction with the infants and toddlers—not so much sitting, more interaction [with the children]."
- There were fewer but almost exclusively positive comments about the **ERS specialist** (14 positive out of 17 total mentions). Most ERS training comments pertained to the TA provider helping with classroom rearrangements and thereby helping programs to secure higher Stars ratings. For example, a director told us, "Yes, [for the ERS specialist] they took pictures of our center, materials, room arrangements and took back the pictures to the supervisors in order to get to the next level."
- The **child assessment training** was mentioned 12 times, and all but one mention was positive. For example, "[B]efore the training, I would've never thought to assess the child, but now I know where to start and the parents are all interested in the progress that the children have taken."
- The exception was **health and nutrition specialty TA**. Directors of seven programs specifically mentioned health and nutrition TA, and those comments were almost evenly divided among strongly positive or strongly negative views about its value. One director said it was the "best" of the six specialty supports, reporting that the other specialty offerings were merely "so-so." A second director said the training "helped my girls realize they were overthinking what they had to do. She was very encouraging." But a third director said it was "totally unhelpful. [The] health and nutrition lady gave kids sugary snacks." And a fourth director called it "a disaster. She didn't know what she was talking about. It was so unhelpful that even my TA [provider] was surprised. The information was either just wrong or contrary to licensing."

In summary, directors overwhelmingly supported five out of the six forms of specialty TA or training. Directors' comments implied that specialty supports were highly applied and yielded tangible benefits for caregivers and teachers, for classroom practices, for children, or for parents.

High-Needs Programs and Stars Plus (Question 6)

When Delaware received its RTT-ELC grant in 2012, it created a cohort of Stars Plus programs.[29] As noted in Chapter 1, Stars Plus is a cohort of programs serving a large proportion of children with POC subsidies that receives weekly rather than monthly on-site visits from their TA provider, attends professional development as a group, and qualifies for special grants tied to each program's QIP. As documented in our Year 2 report, as of October 2014, 145 providers, or approximately 32 percent of Delaware Stars programs, were in a Stars Plus cohort. Small FCCs and centers made up the majority of Stars Plus providers.

Although we did not undertake a formal evaluation of the Stars Plus cohort model, we included an open-ended question in the director interview for Stars Plus providers regarding their experience with TA:

- Do you think the extra support provided for being in the Stars Plus Cohort has helped improve the quality of your program? Why or why not?

Stars Plus Participants Liked the Extra TA Time and the Peer Cohort Structure

We interviewed directors of 63 programs who reported belonging to a Stars Plus cohort (26 percent of programs) as of the time of the interview, either fall 2014 or spring 2015. All but two answered "yes" to our question on whether Stars Plus improved the quality of their program. (The two critical comments were that the level of support was too basic, as were the quality criteria of Delaware Stars generally, and not well organized.) When answering why or why not the Stars Plus supports helped to improve quality, the most common answer related to the TA provider as a form of support—an outside voice or extra set of eyes to advise the director. For example, one director said, "You get more support, more opinions, and more directions to more resources that can help families here." The cited benefits of a dedicated TA provider were fairly global—e.g., a quick go-to resource, a backboard off which to bounce ideas, a trusted advisor who could bring in new ideas and alert directors to potential resources.

Beyond these broad benefits, the other two most frequently mentioned benefits of Stars Plus related to compliance with program regulations and with access to peers via Stars Plus cohorts. As an example of compliance-related benefits: "We know about any updates to Delaware Stars and any new regulations that we need to meet." Belonging to a cohort of programs, meanwhile, was a way to efficiently share good ideas and resources and also encouraged programs to take the steps needed to improve quality: "As a family child care provider, I was alone in this. And

[29] A pilot of the Stars Plus program had been conducted in Wilmington in 2011.

now being in cohorts, I'm not alone anymore." Another director noted, "You're able to meet other providers who are going through similar assessments to get notes and encourage other people as they go through the processes. The classes are much easier when we can work together and share our findings."

Financial Incentives Within Delaware Stars (Question 7)

As noted in Chapter 1, providers that participate in Delaware Stars are eligible for several financial incentives once they have advanced beyond the Starting with Stars rating level. Drawing on Delaware Stars administrative data, we analyzed data for four of the five main types of financial incentives in the Year 2 report: QI Grants; Infrastructure Fund grants; POC Tiered Reimbursement Bonuses; and Compensation, Retention, and Education (CORE) awards. (Data on Teacher Education and Compensation Helps [T.E.A.C.H.] Early Childhood® scholarships were not available.) This analysis documented that these four programs distributed $15.3 million to providers or their staff between October 2013 and September 2014, with the largest share coming from POC Tiered Reimbursement Bonuses (61 percent, or $9.4 million). On average, Delaware Stars programs received approximately $27,000 per program in combined incentives. The highest participation rate, measured as the share of providers participating in a financial incentive during the year, was for QI Grants (55 percent of programs), followed by POC Tiered Reimbursement Bonuses (45 percent), CORE awards (28 percent), Infrastructure Fund technology awards (27 percent), and Infrastructure Fund capital awards (9 percent). These patterns reflect, in part, the differential eligibility for each incentive depending on the Stars level.

The director interviews allowed us to gain further understanding about how directors of participating programs in Delaware Stars view these financial incentives and the impact they have on program quality.

Directors Value the Financial Incentives, Especially the Tiered POC Reimbursements

We asked directors whether their program had ever received the three types of financial incentives that apply at the provider level—POC Tiered Reimbursement Bonuses, Infrastructure Fund grants, and QI Grants—and whether one or more staff had ever received the two types of incentives that are awarded to staff—CORE awards and T.E.A.C.H. scholarships. In addition, we asked directors to pick which incentive, of the five, they would rank as the most important. For each type of financial incentive, Table 5.2 shows the percentage of programs receiving it and the percentage where the director rated it as most important. The financial incentives are listed from most to least important based on the directors' rankings.

The most common financial incentive received was a QI Grant, with a recipiency rate of 61 percent according to director responses, but just 13 percent rated this financial incentive as most important. POC Tiered Reimbursement Bonuses were the second-most-common incentive

Table 5.2. Responses by Directors of Programs in Delaware Stars Regarding Importance of Financial Incentives

Type of Financial incentive	Weighted Percentage Ever Receiving	Weighted Percentage Ranking Most Important
Awards to providers		
POC Tiered Reimbursement Bonuses	59.2	43.2
QI Grants	60.8	12.6
Infrastructure Fund grants	49.9	12.0
Awards to staff		
CORE awards	47.0	11.4
T.E.A.C.H. scholarships	40.2	10.3

SOURCE: Authors' analysis of provider sample director interview responses.
NOTES: Based on 154 respondents. Results are weighted to account for provider sampling and nonresponse. For the ranking question, a residual 18 percent of directors answered "other," with the modal "other" answer that all types of grants were important or some combination of two or three of them.

received, applying to 59 percent of programs.[30] Most notably, directors were most likely to rate this incentive as most important (43 percent) compared with the four other incentives. This may not be surprising, given that our analysis of administrative data showed this financial incentive to have the largest dollar impact. The other three incentives have also reached, at some point, 40 to 50 percent of all Delaware Stars programs, according to directors' reports.[31] Only a minority of directors (from 10 percent to 13 percent) viewed the two grant award programs and the awards to staff as most important among the five incentive programs.

If a director indicated receiving a particular type of grant or subsidy, we then asked whether that grant or subsidy improved the quality of the program and why or why not. We begin by summarizing responses regarding the three types of financial incentives that are received by programs—POC Tiered Reimbursement Bonuses, QI Grants, and Infrastructure Fund grants—and conclude with CORE awards, which are awarded to individual staff. (We did not ask for more details on directors' views regarding T.E.A.C.H. scholarships.)

POC Tiered Reimbursements

Providers with a Star 3, Star 4, or Star 5 rating get 80 percent, 93 percent, or 102 percent of the 2011 market rate, respectively, for children who are eligible for POC, an increase over the usual reimbursement rate. The funds are distributed on a monthly basis. As indicated in Table 5.2, about two-thirds of directors indicated that their program had ever received POC Tiered

[30] This rate exceeds our estimate of 45 percent for the 2013–2014 period covered by the administrative data analyzed in the Year 2 report. However, directors reported "ever receipt," which would be expected to be higher than receipt in a given year because whether a provider has one or more enrolled children receiving POC subsidies can vary from year to year.

[31] Again, the participation rates in Table 5.2 exceed our one-year participation rate estimates, likely because programs or their staff may receive the incentives some years and not others.

Reimbursement Bonuses. Of these, the large majority described the funds as substantially helpful when asked whether this was an effective tool to improve program quality. One director noted, "Yes, our center would not be open without it. When Obamacare came into effect, without tiered POC reimbursement we would not have been able to sustain the costs of healthcare for the staff and keep the center open." A second director told us that she was able to offer a 4-percent match on a 401(k) retirement plan as a result of the funds, as well as health insurance for her employees. Many described POC tiered reimbursements as useful, but as top-off funds to defray smaller costs, such as replacing materials that are in high use and frequently destroyed, purchasing supplies, or providing smaller cash incentives to the teachers so they feel more appreciated. Two directors noted a lack of accountability in the way that tiered dollars are spent once received by programs, noting that these added funds should be required to be spent on program operations or POC children. Several respondents noted that the POC tiered reimbursements are appreciated but, even so, did not come close to covering the actual costs of running an ECE program. One director noted that the POC tiered reimbursement rate was only "$20 for 9 hours of child care and they expect you to be grinning and jumping but it's not fair."

QI Grants

QI Grants are for materials and professional development, tied to the QIP, for programs at Star 2 to Star 4 and vary with program type and size. The grants range from a maximum of $750 for small FCC providers up to a maximum of $5,000 for the largest centers. All but one of the 81 directors who indicated receiving a QI Grant felt that it had improved the quality of the program. All had used the grant funds to buy various materials for the program, and many directors noted that it had helped their programs move up a star level, consistent with the intent of the grant.

Infrastructure Fund

This fund provides grants to programs at Star 2 or higher that enroll children with POC subsidies. Improvements may be for capital or technology but need to be related to the QIP. Directors of 73 programs indicated receiving infrastructure funds for capital improvements or technology needs. Twenty indicated using the funds to buy computers or iPads for teachers or classrooms or themselves. Other common uses were to install such physical items such as a fence or sinks in rooms, build or move an interior wall, or invest in the playground. Those who received the grants were invariably positive about it. Five directors noted that their applications were rejected, and several of these directors noted inconsistent guidance and stated that they would not have applied had they known that they were ineligible because they did not serve POC children.

CORE Awards

CORE awards are for professionals who meet specified educational requirements. Directors of 88 programs indicated that one or more of their staff have or had received a CORE award. Of

these directors, 80 percent responded that CORE awards were useful for improving the quality of the program. Many noted that, even if these were only one-time awards, they helped retention and morale. But several directors mentioned that the rules changed over time, lowering the award amounts, without good communication, disappointing applicants. The 20 percent who indicated that the awards did not improve program quality wished that staff had been recognized in other ways and that the funds could be redirected to grants for programs rather than staff.

Consumers and Delaware Stars (Question 8)

As part of our Year 1 report (Schwartz et al., 2014), we gathered information from interviews and focus groups to understand stakeholders' experiences with Delaware Stars. One finding that emerged from those discussions was the perception that the general public's awareness of Delaware Stars was low and that understanding of the quality ratings was lower still. In focus groups, for example, parents indicated a desire to understand how the Delaware Stars standards influenced or related to ECE program quality, yet the available outreach approaches and promotional materials as of that time appeared to be insufficient to help families understand the quality rating system.

To explore these issues further, we included questions in the director interview about the outreach methods that programs use to market their Delaware Stars participation and directors' perceptions about whether Delaware Stars ratings influence parental choice.

Directors Report Low Parental Recognition and Low Marketing Value from Participation in Delaware Stars

Table 5.3 reports on responses from directors in programs participating in Delaware Stars regarding communication with the public about their ratings. Among the directors of programs in Delaware Stars, a large majority (80 percent) reported informing parents about their Delaware Stars rating. Directors who reported informing parents indicated that they used multiple strategies for doing so. Nearly 79 percent of directors reported using verbal communication, and nearly as many also said they displayed posters, signs, or banners (74 percent). Other common forms of communication included distribution of prepared materials from the state about Delaware Stars in their applicant packets (51 percent) and referring parents to the Delaware Stars website (38 percent). Among the small number of directors that participated in Delaware Stars but reported that they did not advertise to parents, the main explanations were that they did not see the need; they were new to the program; or they found that the information was not useful to parents (i.e., other criteria mattered more).

Table 5.3. Responses by Directors of Programs in Delaware Stars Regarding Public Communication

Survey Item	Weighted Percentage of Respondents
Director informs prospective parents about the program's Delaware Stars rating	80.2
Among those providing information, methods used to inform prospective parents about the rating	
Explain verbally to applicants	79.2
Hang up posters, signs, or banners in the program	74.4
Include printed outreach materials provided by OEL in application packet	50.5
Refer applicants to Stars website	37.8
Applicants ask if program is rated by Delaware Stars	
Yes	32.3
No, applicants rarely, if ever, ask	54.8
Other	12.9
Program receives more applicants because of the Delaware Stars rating	
Yes	30.7
No, applicants rarely, if ever, ask	69.3

SOURCE: Authors' analysis of provider sample director interview responses.
NOTES: Based on 162 respondents. Results are weighted to account for provider sampling and nonresponse.

At the same time, a majority of directors (55 percent) indicated that parents rarely or never asked about Delaware Stars (see Table 5.3). Among directors who said that parents did ask, they estimated, on average, that about one in five parents asked about the program's rating. Given this low rate of parental awareness, it is not surprising that just one-third of directors (31 percent) indicated that they received more applicants to their program because of their participation in Delaware Stars.

Based on our analysis of open-ended responses related to parents' perceptions, about three-quarters of directors (73 percent) thought that parents perceived Delaware Stars to be a good gauge of quality, but almost all directors stressed that only those parents who knew to ask about it or were educated by staff about it viewed it as a good gauge. Directors were almost universal in commenting that parental awareness of Delaware Stars was low. At the same time, as discussed later in this chapter, a majority of directors reported participating in Delaware Stars to make their program more attractive to parents. Further, a majority of directors viewed recognition from parents, other providers, and the public that their program offers high-quality care as a benefit from participating in Delaware Stars.

Other Aspects of Delaware Stars Systems Performance (Question 9)

The first two evaluation reports used Delaware Stars administrative data to provide a detailed portrait of participation in Delaware Stars and rating outcomes for participating providers. As of October 2014, we found that licensed centers had the highest participation rate (73 percent), followed by large FCCs (38 percent), and small FCCs (23 percent). Data from OEL as of May

2016 document a continued increase in the Delaware Stars participation rate, with centers climbing to 79 percent, large FCCs attaining 51 percent, and small FCCs reaching 27 percent (Delaware Institute for Excellence in Early Childhood, 2016). This high participation rate for centers reflects the initial focus of Delaware Stars on recruiting center-based programs into the system and the subsequent focus on enrolling FCCs. These participation rates are similar to other statewide QRISs that are at a comparable stage of implementation (BUILD Initiative, 2016). Because center-based programs enroll the largest number of infants, toddlers, and preschool-age children, the high participation rates for centers in Delaware Stars has meant that the system is reaching a large share of children in licensed-based settings. In particular, our estimates indicate that 66 percent of children in licensed settings as of October 2014 were in programs participating in Delaware Stars. This means that the quality improvement aspects of Delaware Stars have a large reach.

The administrative data also confirmed that programs participating in Delaware Stars were moving upward through the rating tiers. As of October 2014, 45 percent of Delaware Stars programs were at the top two rating levels: Star 4 or Star 5. Among children enrolled in a Delaware Stars program, 62 percent were in a Star 4 or Star 5 program. In addition, over time, more programs were reaching Star 3 and above through the regular rating process. As of January 2014, more than half of Star 4 providers (58 percent) and more than one-third of Star 5 providers (39 percent) were stand-alone Head Start or ECAP programs or NAEYC-accredited programs, which did not require the regular rating process. By October 2014, these shares fell to 39 percent and 20 percent, respectively, indicating that more programs not in the alternative rating pathway were moving up to the higher levels.

The director interview provided an opportunity to explore topics—beyond those that could be considered with the administrative data—that are relevant for understanding experience with the rating structures and processes, as well as overall views on the reasons for participating in Delaware Stars.

Experience with Rating Structure and Process

Both the Year 1 and Year 2 reports examined the timing of transitions from one rating tier to the next based on Delaware Stars administrative data. Those analyses showed that programs transition most rapidly out of Starting with Stars to Star 2 and from Star 3 to Star 4. Some providers at Star 2 move quickly to Star 3, but a substantial share of programs remain at that level for 18 months or more. Likewise, the transition from Star 4 to Star 5 is a more gradual process (Schwartz et al., 2014; Auger, Karoly, and Schwartz, 2015).

Time and Resources Are Top Barriers to Moving Up in Stars

The director survey included several questions that offer additional perspective on providers' experience with the rating structure and process. In particular, we asked directors whether they

were actively planning to move up to the next star level. Among those not already a Star 5 program, almost all indicated that they were working toward a higher rating (see Table 5.4).

We then posed a list of nine potential barriers to moving up in the Delaware Stars ratings (plus a tenth "other" option) and asked which obstacles they faced in moving up to the next star level and to rank their selections from most important (ranking of 1) to least important. Table 5.5 presents the list of barriers and indicates the percentage of programs where the director indicated the barrier was an obstacle and the average rank for each barrier among those that selected it. The barriers are listed from highest to lowest prevalence.

Overall, the most commonly selected barrier was the need to get paperwork and documentation in order (55 percent), followed closely by finding time to complete the tasks for the next level (52 percent). The other barriers were not cited as often, from 37 percent of program directors to 12 percent. Among the barriers mentioned, the highest average rating was for finding the time to complete the tasks (average rank of 2.0, where a rank of 1 was allocated for the most important barrier), followed by an average rank of 2.3 for having insufficient funding to meet the standards. Thus, resources and time appear to be a common theme among the most prevalent barriers.

Directors Viewed Several of the Essential Standards as Being Among the Most Difficult to Achieve

A key feature of Delaware Stars is that it provides program directors partial choice to pick the standards on which they are validated to reach Star 3 and above. For example, there are 34 total standards from which to choose for ECE centers. So long as programs surpass both a minimum number of points in each of four quality domains and a minimum score on the ERS, they can reach Star 3 to Star 5. Starting in 2015, however, the state for the first time made some standards mandatory to reach certain star levels. We listed these essential standards in Chapter 1 and analyzed their relationships to program quality and child outcomes in Chapters 3 and 4, respectively.

Table 5.4. Responses by Directors of Programs in Delaware Stars Regarding Plans to Move Up in Delaware Stars

Survey Item	Weighted Percentage of Respondents
Director is actively planning to move up to the next star level	
Yes; I have applied or I am planning to apply to move up another level	75.0
No; I have no plans to apply to move up another Star level	1.3
Not applicable (already rated as Star 5; no higher level)	22.1
Other	1.6

SOURCE: Authors' analysis of provider sample director interview responses.
NOTES: Based on 154 respondents. Results are weighted to account for provider sampling and nonresponse.

Table 5.5. Responses by Directors of Programs in Delaware Stars Regarding Barriers to Moving Up in Delaware Stars

Barrier	Percentage Selecting Barrier	Average Ranking for Barrier When Selected (1 = Top Barrier)
Getting the paperwork and documentation in order	55.2	2.5
Finding time to complete tasks required for the next level	51.9	2.0
Preparing for and meeting required ERS score	36.7	3.4
Completion of required staff education and training	34.7	2.5
Insufficient funding to meet standards	34.2	2.3
Challenges in developing a curriculum	25.5	3.6
Lack of time or resources due to preparing for and meeting national accreditation standards	24.8	4.0
Having to wait six months to get assessment	17.8	3.9
Needing more feedback from TA provider	12.2	3.8
No obstacles	3.5	–

SOURCE: Authors' analysis of provider sample director interview responses.
NOTES: Based on 156 respondents. Results are weighted to account for provider sampling and nonresponse.
– = not applicable.

Directors named some of these essential standards as the hardest ones to meet. Directors of 71 programs answered our question "Are there particular standards in the Delaware Stars system that you view as especially difficult to attain? If so, which ones and why?" The most frequently mentioned difficulties related to the Staff Qualifications domain. Directors of 23 programs mentioned challenges here, ranging from the cost of obtaining a Delaware Administrator Credential (which is an essential standard), to getting staff to obtain Delaware credentials, to achievement of certain steps on the Delaware Early Childhood Career Lattice. The next-most-cited challenge related to a written curriculum, which is another essential standard. Directors at ten programs mentioned the requirement to have a written curriculum as especially challenging—whether getting staff to use the curriculum; ensuring that all students had handouts and sheets; or, for those writing their own curriculum, how to do so. Directors at eight programs called out playground requirements for specific equipment, setup, or ground covering as being the most difficult, because of cost (e.g., "They want you to have like 6 to 9 inches of mulch, do you know how expensive that is?"). Finally, directors at six programs called out the following standards as being the most difficult: individual assessments of children (an essential standard), establishing community partnerships because of the daytime hours required to do so, and getting sufficient parental involvement to complete surveys and return forms.

We also posed the question "Are there particular standards in the Delaware Stars system that you view as especially easy to attain? If so, which ones and why?" Directors of 155 programs answered the question, typically referencing things as easy that they were already doing or had been doing before applying to Delaware Stars. The most frequently mentioned standards (29 mentions) belonged to the Family and Community Partnership domain—such standards as

providing regular written correspondence with families, having procedures that welcome all families and children, and implementing family-centered events annually. The other standards mentioned as being easy were straightforward ones like having certain classroom materials (for the ERS, 14 mentions); employing staff with certain qualifications, which is easy for those whose staff already have them (14 mentions); and type of teacher-child interactions as rated in the ERS (nine mentions).

Overall Experiences with Delaware Stars Participation

Several questions in the director interview took a broader perspective in asking about the factors motivating participation in Delaware Stars and how participating directors see the benefits from participation.

Improving Quality Was a Top Motivator for Participation in Delaware Stars

We asked directors to select from a list of eight possible reasons for being in Delaware Stars. As seen in Table 5.6, the top-ranked reason (and the one that 85 percent of respondents chose) was to improve the quality of their program (average ranking of 1.9). Three other explanations were almost as prevalent: access to the grants and other financial incentives (72 percent), making the program more attractive to parents (64 percent), and wanting new ideas for their program (62 percent). The least-prevalent response and the one with the lowest average rank was to attract and retain qualified staff (41 percent).

Table 5.6. Responses by Directors of Programs in Delaware Stars Regarding Reasons for Participating in Delaware Stars

Reason	Weighted Percentage Selecting Reason	Weighted Average Ranking for Reason When Selected (1 = Top Reason)
I wanted to improve the quality of my program	85.3	1.9
I wanted access to the grants and other financial incentives available via Delaware Stars	72.2	2.7
I wanted to make my child care or preschool program more attractive to parents	64.3	3.9
I wanted new ideas for my child care or preschool program	62.0	4.5
I wanted to increase my business	56.8	5.1
I wanted the technical assistance that Delaware Stars offers	54.1	4.4
I wanted more professional recognition	51.0	3.7
I wanted to attract and retain qualified staff	40.5	4.7

SOURCE: Authors' analysis of provider sample director interview responses.
NOTES: Based on 154 respondents. Results are weighted to account for provider sampling and nonresponse.

We also listed six aspects of Delaware Stars—TA, grants, recognition, ERS scores, validation process, and marketing—and asked directors which were the most beneficial (see Table 5.7). TA was the most prevalent feature selected as beneficial (81 percent) and was also the highest ranked of the six aspects (average rank of 1.7). Grants and financial incentives were a close second in terms of being selected (74 percent) and the assigned rank (average rank of 2.3). The value of Delaware Stars participation as a marketing tool was the least likely to be selected (41 percent) and also the lowest ranked (average rank of 4.0).

In addition, we asked directors who were not in Delaware Stars why they were not in the program. Because we had only a few such providers in the provider sample, just ten directors answered the question. However, among these ten, no one theme dominated. For example, one director answered that it took too much time to enroll. Two answered that Delaware Stars was not a good gauge of quality. Others had idiosyncratic answers, such as being on probation (and thus not being allowed to enroll), being a part-time program, or having a plan to rejoin after being dropped.

Most Directors See Multiple Benefits for Their Staff from Delaware Stars

For those directors in Delaware Stars, we closed the spring 2015 interview with the question "Can you talk about the impact Delaware Stars has had, if any, on your staff—particularly teachers? Have you seen a change in your teachers? If yes, how so? If not, why do you think there is no change?" Directors of 153 programs answered the question.

Three quarters of directors (111 out of 153, or 73 percent) responded that Delaware Stars had helped their programs raise the bar for teachers' practice. Directors frequently mentioned that the program helped to professionalize the teachers, gave them a structure of what to aim for and do,

Table 5.7. Responses by Directors of Programs in Delaware Stars Regarding Beneficial Aspects of Participation in Delaware Stars

Benefit	Weighted Percentage Selecting Benefit	Weighted Average Ranking for Benefit When Selected (1 = Top Benefit)
TA	80.7	1.7
Grants and financial incentives	73.7	2.3
Recognition I get from parents, other providers, or public that I am providing high quality care	59.5	3.0
Environmental Rating Scores for my classrooms	53.9	3.0
Validation process (including rating on individual standards and the ERS)	43.6	3.6
Participation provides me with a marketing tool for my child care or preschool program	40.6	4.0

SOURCE: Authors' analysis of provider sample director survey responses.
NOTES: Based on 156 respondents. Results are weighted to account for provider sampling and nonresponse.
– = not applicable.

validated teachers in programs at higher star levels, motivated teachers to try harder, and helped teachers understand better why they should teach in certain ways. For example, one director answered: “I've seen them grow into intentional educators. It gives them direction and assistance in guiding them because it shows them that we’re following the Delaware Stars standards and not my standard.” Another said, “Pre-stars they didn’t understand why the job was so important. Now they have a deeper understanding of roles and why they do what they do.”

Many directors indicated that participating in Delaware Stars had a profound impact on their staff. “When we first got into Stars, the staff didn’t like it and they all quit,” said one director. “I replaced the staff, and I find that most of the staff are liking it, with a few stragglers.” While many noted that being observed and going through change was at least temporarily anxiety-producing, they found, overall, that the experience was worth it. At the same time, a small minority of program leaders (20 directors) indicated that stress was teachers’ primary experience with Delaware Stars. “It has made them miserable, Delaware Stars has made them hate the field and make them leave the field because a lot of the work is expected without an increase in pay.” Another answered, “Honestly it was a very stressful year with Stars. It seems like we started the process in November and ended in April. It was stressful for the staff, and everyone was miserable all year.”

Conclusion

The multiyear RAND evaluation has addressed aspects of the QRIS related to financial incentives, quality improvement, and other aspects of system performance that complement the validation analyses reported in Chapters 3 and 4 and provide a portrait of the QRIS as a whole. Data collected earlier in the evaluation through interviews and focus groups, two rounds of analyses of administrative data from Delaware Stars, and responses to the director interview conducted as part of the 2014–2015 data collection all help to inform important questions related to TA, financial incentives, the role of Delaware Stars in parental provider choice, and other aspects of system operations. Although we discussed these as separate topics in this chapter, they are interrelated in ways that we highlight in this concluding section.

TA supports and financial incentives are designed to support programs through the process of quality improvement and to sustain them at the higher levels of quality that they achieve. Administrative data document high participation rates in TA, at a frequency consistent with the planned level of support. Delaware Stars system data also document the considerable investment made in financial incentives, both at the system level and in terms of the combined value of the incentives on a per-program basis. In survey responses, directors were strongly enthusiastic about the TA supports within Delaware Stars. They valued TA—especially specialty TA. The directors referred to TAs as navigators who helped programs wade through paperwork; complex and often changing program regulations; and multi-stage, multi-department processes. Directors also valued financial supports offered through Delaware Stars, most particularly the POC Tiered

Reimbursement Bonuses, the largest component of the financial incentives. The potential importance of these financial incentives is all the more apparent, since directors indicated insufficient funding to meet standards as a top barrier to moving up in star levels. Those in Stars Plus also endorsed the supports of having a peer network of other directors with whom to share ideas and lessons.

Although TA and financial supports were the top-ranked benefits of Delaware Stars, directors also indicated three aspects of TA in which there was room for improvement. The first related to greater consistency in guidance among the TA providers themselves and also between the TA providers and other OCCL or Delaware Stars staff that directors came into contact with, such as assessors, validators, and OCCL staff. Directors also typically wanted more time with their TA providers. Finally, several directors were critical of the degree of turnover among TA providers caused by reassignments.

While most directors in Delaware Stars reported supporting the vision of quality that Delaware Stars promotes, some also indicated that several of the most challenging standards to meet are the very ones that Delaware Stars is newly mandating to reach Star 4 and Star 5. Staff credentials, curriculum, and child assessments were viewed as among the hardest standards to meet. The easiest, by contrast, were believed to be the ones within the Family and Community Supports domain that related to outreach to parents that most directors claimed they were already doing.

An issue identified in the first year of the evaluation was parents' limited understanding of Delaware Stars. From the perspective of directors, this appears to be an area where further improvement may still be needed. Most directors indicated that they were providing information about their participation in Delaware Stars and their rating, but only about a third of those directors reported that parents inquired about ratings or thought that the ratings influenced parental decisionmaking. Even though a majority of providers wanted to improve quality to attract more families, most providers did not think that they were benefiting from Delaware Stars in terms of their marketability and enrollments.

Chapter 6. Key Findings and Recommendations

States and localities implementing ECE QRISs have been strongly encouraged to validate their systems, and a majority of them have completed such studies or have one in process. Key aims of these analyses include understanding whether their rating scales are capturing meaningful differences in program quality and assessing the functioning of other aspects of the system, particularly those components focused on quality improvement, the "I" in QRIS (Zelman and Fiene, 2012). The requirement for a QRIS evaluation as part of the federal RTT-ELC grant mechanism provided further impetus for such evaluation studies. This evaluation of Delaware Stars was undertaken as part of the state's RTT-ELC funding.

A key objective of this evaluation was to examine whether the Delaware Stars rating system captures meaningful differences in program quality. Based on primary data we collected in 2014–2015, we analyzed the relationship between Delaware Stars ratings and both alternative measures of program quality and measures of children's learning and development. Over the course of the evaluation, administrative data and responses to the director interview also implemented in 2014–2015 provide further perspective on the performance of Delaware Stars as a QRIS. In this final chapter, we summarize our key findings with respect to each of the study questions, place those findings in the context of research on QRISs more generally, and offer recommendations for stakeholders in Delaware regarding the QRIS.

Summary of Key Findings

Table 6.1 summarizes our research questions and key findings.

Table 6.1. Study Research Questions and Key Findings

Study Research Questions	What We Found
Delaware Stars Ratings and Program Quality	
Q1. To what extent do the quality tiers of Delaware Stars accurately reflect differential levels of program quality such that programs at the top levels provide a higher-quality care and early learning experience than programs at lower levels?	• Alternative quality measures rose modestly with Delaware Stars ratings, but the increases were usually statistically insignificant and small in magnitude. • For the most part, alternative quality measures were not positively related to scores on the components that make up the overall Delaware Stars rating (i.e., ERS and four quality domains in which points are obtained). • There were no statistically significant relationships between the essential standards and alternative quality measures.
Q2. What is the relationship between program characteristics and quality in Delaware Stars?	• FCCs had high quality scores on some measures, but this may reflect sample selectivity. • Programs serving majority ECAP children had lower instructional support quality, on average.

Table 6.1. Study Research Questions and Key Findings, Continued

Study Research Questions	What We Found
Delaware Stars Ratings and Child Developmental Outcomes	
Q3. All else being equal, do young children participating in higher-rated programs have better learning and developmental outcomes than similar children in nonparticipating or lower-rated programs?	• Differences in children's development across rating tiers were generally small and statistically insignificant. • Children in Star 5 center-based programs modestly outperformed children in Starting with Stars and Star 2 programs on executive function skills, an important predictor of academic success. • There was no evidence that top-rated Delaware Stars center-based programs conferred greater benefits for children from low-income families as compared with lower-rated programs.
Q4. What dimensions of Delaware Stars program ratings are most vital to child learning and developmental outcomes?	• For center-based programs, points obtained in two quality domains—Management and Administration and Qualifications and Professional Development—were associated with higher scores on some assessments of early academic and cognitive skills. • Children in center-based programs that met more of the six essential standards modestly outperformed children in programs that met fewer essential standards in terms of executive function skills.
Delaware Stars System and Program Quality Improvement	
Q5. To what extent does the TA (i.e., on-site support, orientation, etc.) provided to Delaware Stars participants help providers to move up in Delaware Stars?	• TA was the second-largest financial investment made by Delaware Stars (after POC Tiered Reimbursement Bonuses). • Delaware Stars programs received the intended number of TA visits. • Directors highly valued TA overall, especially the specialty TA and specialized training designed to improve program quality in targeted areas. • Directors identified the need for greater consistency in TA guidance, more frequent TA visits, and less frequent reassignments of TA staff.
Q6. To what extent do high-need programs that participate in Stars Plus enhance their program quality?	• Directors in Stars Plus viewed the peer network and the extra TA they received as valuable supports for improving quality.
Delaware Stars System Performance	
Q7. Are the financial incentives and supports for providers sufficient to support the needed quality improvements?	• The five types of financial incentives represent a large investment overall and conferred an average of over $27,000 per program in 2013–2014. • Directors value these incentives as supports for making and sustaining program quality improvements, especially the POC Tiered Reimbursement Bonuses, which were the largest incentive measured by total dollars as of 2013–2014.
Q8. What do consumers understand about Delaware Stars? Do consumers ask about ratings?	• Directors reported using Delaware Stars marketing materials but also noted low parental awareness of Delaware Stars and saw low marketing value for the program.
Q9. How well do the Delaware Stars system components operate?	• Administrative data show continued recruitment of programs into Delaware Stars and upward movement among those in the system, with especially high participation rates among centers. • Time and resource constraints were the most-cited barriers to moving up the Delaware Stars rating tiers. • Directors viewed several of the essential standards as the most difficult to attain. • Improving quality is a top motivator for participating in Delaware Stars. • Directors identified multiple benefits for their staff from being in Delaware Stars.

Delaware Stars Ratings, Program Quality, and Child Developmental Outcomes

For the most part, based on data from the provider sample and child sample, we did not find the expected large and consistent positive differences by Delaware Stars ratings in our alternative measures of program quality or measures of child developmental outcomes. Our alternative measures of quality captured both structural features (e.g., dimensions of the learning environment and staff qualifications measured in the PQA) and process components (e.g., dimensions of teacher-child interactions measured in the PQA, CLASS, and CIS). Our measures of child development included assessments of early academic skills, such as language, literacy, and mathematics; a measure of executive function; and measures of social-emotional and behavioral skills.

Although we found in several instances the expected positive relationship between Delaware Stars ratings and program quality, the incremental improvements in the alternative measure of quality tended to be very small for each step increase in the Delaware Stars rating level, especially in moving from Star 3 to Star 4 to Star 5. Effectively, for providers within each Delaware Stars rating level, there was considerable variation in quality according to the PQA, CLASS, and CIS. We also considered the relationship between the components that are used to derive the Delaware Stars rating and the alternative measures of quality. In almost all cases, there was no strong positive relationship between the alternative measures of quality and the ERS score, the four quality domains in which points are obtained, or the essential standards.

Likewise, there were some child outcomes that showed the expected stair-step pattern in moving from Star 3 to Star 5 ratings, but the differences were small in magnitude and never statistically significant. The one significant difference across rating tiers was for the measure of executive function in contrasting the two lowest tiers in Delaware Stars with the highest tier. Selected academic measures and the measure of executive function also showed the expected relationship with some of the quality domains or the essential standards, but the patterns were not consistent.

Delaware Stars System Performance

Our evaluation considered other aspects of Delaware Stars beyond the structure of the ratings. Overall, ECE program directors are motivated to be in Delaware Stars in order to improve their quality. Indeed, Delaware Stars makes considerable investments in quality improvement—the "I" part of the QRIS. This is important because directors report that it is challenging to advance program quality and barriers in terms of time, and financial resources can get in the way. In terms of TA, administrative data documented high participation rates. Surveyed directors expressed generally positive views about TA, especially specialty TA, although they identified scope for improving TA consistency, frequency, and constancy (through reduced TA provider turnover at the program level). Financial incentives were also assessed by directors as important for supporting quality improvement and sustaining programs at a higher level of quality.

Although we did not undertake formal polling of parents of young children, earlier focus groups in the first year of the study and program directors both pointed to the limited influence that Delaware Stars ratings appear to have in determining parental choice of ECE programs. Consequently, most providers do not think that they are benefiting from Delaware Stars in terms of their marketability and enrollments.

Limitations of Analysis

As noted in Chapter 1, it is important to recognize the limitations of the analyses undertaken for this study and their implications for the findings. Most importantly, as of the time of the study, Delaware's QRIS had gone through several changes and was continuing to undergo change even as our validation study was in the field. For example, for a majority of the ECE programs in the provider sample, the six essential standards now required for reaching Star 4 and Star 5 had yet to be made mandatory. This means that we did not examine the fully implemented version of Delaware Stars that is now in effect. It is possible that the requirement to meet the essential standards will support a stronger association between Delaware Stars ratings and external measures of program quality, such as the ones we analyzed. And our analyses demonstrated, in one instance, the expected relationship between the bundle of essential standards and child development. On the other hand, the associations were not particularly strong, so it is not clear that the move to a set of required standards will ensure that the rating structure performs as intended.

Other caveats noted earlier are relevant to keep in mind as well. These include that the provider sample primarily consisted of center-based programs (including schools), so we did not have the ability to look for differences in the performance of the QRIS for center-based versus home-based settings. We also could not study the validity of the QRIS for capturing quality differences in the care environment for infants. While selectivity bias because of parental choice of the ECE settings could be a concern, we would expect the bias to make it more likely to find a positive relationship between children's development and program quality, whereas we found only a very weak association, at best.

Finally, it is the case that our analyses are constrained by the set of quality measures we chose to collect and the aspects of children's development we assessed. Given that quality is a multidimensional concept, it may be that other measures of program quality than the ones we collected would show a stronger relationship to Delaware Stars ratings. And there may be other areas of children's development that are more strongly associated with program quality, as summarized in the Delaware Stars rating structure. At the same time, the measures of quality we employed captured both structural and process aspects of quality, and they were derived from an extensive research literature that has documented the importance of these aspects of quality for children's development. The child development assessments included early academic measures of early literacy and math, executive function, and social-emotional and behavioral skills—all dimensions of children's learning that high-quality ECE programs would be expected to affect.

Validation of Delaware Stars in Context

Given the full-scale implementation of Delaware Stars and the significant resources invested in the system, it is important to place the findings, particularly those pertaining to the validation portion of the study, in context. Two frames are most salient for interpreting the findings from our analysis of Delaware Stars: how they compare with those of similar studies for other statewide or local QRISs and how the findings relate to the larger research on the relationship between ECE quality and child developmental outcomes.

Findings for Delaware Stars in Context of QRIS Validation Studies in Other States and Localities

As part of Chapter 1, we provided a summary of the published QRIS validation studies that considered the relationship between QRIS ratings and either external measures of program quality or child developmental outcomes using longitudinal data and controls for child and family background (or both). Viewed from the perspective of the findings from those other studies, our results for Delaware Stars are entirely consistent with what has been found for other QRISs. Although most validation studies find the expected positive relationships between QRIS ratings and alternative measures of program quality, the correlations are generally weak. The increase in the average level of provider quality in moving from rating tier to rating tier is small by comparison to the implied movement embedded in the rating scale (such as the one-scale-point increase in the ERS to move in Delaware Stars from Star 3 to Star 4 or from Star 4 to Star 5). At each rating level, a consistent finding is that, according to the alternative measures, there is considerable variability in quality for programs in the same rating tier. Likewise, the lack of a strong relationship between children's development and QRIS ratings is a common finding. Finally, in many cases, the absolute level of program quality, even at the highest rating tier, has not been at the level of programs with demonstrated impact on children's development.

One issue may be that global measures of ECE program quality, in particular, are not very stable or that there is more variability in quality across classrooms within a given program that rating systems do not account for. As designed, QRISs, for the most part, assume that program quality is relatively stable over time and that program quality—comprising features at the classroom level and program level—can be aggregated to a single center-level measure. Across the 37 QRISs summarized in Table 1.4, for example, the modal time interval for which ratings are valid is three years, the same practice followed by Delaware Stars. If highly rated programs tend to diminish in quality between ratings, whereas the reverse is true for lower-rated programs (because they are trying to improve), we might expect to see a low correlation between alternative quality measures at a point in time and the stated rating tiers based on earlier quality measurement.

Other sources of measurement error could include a lack of inter-rater reliability on the part of the QRIS verification staff or on the part of the field staff used by the independent evaluator

(or both). Observation staff for Delaware Stars are required to reach at least 85 percent agreement with a master coder to achieve reliability on the ERS and are reassessed periodically to ensure that they remain reliable (BUILD Initiative, 2016). The RAND field staff likewise followed reliability standards established by the developers of the PQA and CLASS that required at least 80 percent agreement prior to starting fieldwork and during checks while fieldwork was under way (see Appendix B). In both cases, the reliability standards do not eliminate errors in quality measurement using tools like the ERS or those used in this study (Karoly, Zellman, and Perlman, 2013). Another possible source of measurement error arises when quality is not uniform across the classrooms in a given program. As discussed in Karoly, Zellman, and Perlman (2013), empirical evidence suggests a nontrivial degree of variability across classrooms in their measured quality using tools like the ERS. Indeed, as discussed in Appendix G, this was also the case for Delaware Stars programs in our provider sample when more than one classroom was observed. On average, PQA scores across the lowest- and highest-scoring classrooms in the same program differed by 0.6 scale points, CLASS scores differed by 0.9 to 1.3 scale points, and CIS scores differed by 0.5 points. This within-center variability across classrooms has implications for program ratings, especially when not all classrooms are measured to calculate the program rating, as is the case with most QRISs, or for purposes of a validation study, such as this one.

It is also important to emphasize that QRIS validation studies, like our study of Delaware Stars, do not provide an evaluation of the impact of the QRIS on child outcomes. Our empirical findings indicate that we are not detecting a strong positive relationship between Delaware Stars quality ratings and children's learning and developmental outcomes. That is not the same thing as saying that there have been no improvements in child outcomes as a result of implementing Delaware Stars. Even if the Delaware Stars rating structure does not differentiate program quality strongly, if (1) ECE program quality is improving over time in ways that favorably affect children's development and (2) that improvement is happening for much of the ECE provider base, we would expect that at least some children will have experienced better developmental outcomes as a result of the implementation of Delaware Stars compared with a status quo where the QRIS had not been implemented. In other words, it is the quality improvement process itself and the diminishing share of lower-quality providers in the marketplace—either because they improve their quality or go out of business for lack of demand—that could drive the beneficial impact of Delaware Stars on children's outcomes.

Findings for Delaware Stars in Context of Research on Quality and Child Outcomes

It is also important to recognize that QRISs have gained currency as a mechanism for ECE quality verification and program improvement at a time when evidence is accumulating that the available measures of quality, either global quality measures or specific indicators, may not be as strongly related to children's outcomes as suggested by earlier research. A study by Sabol et al. (2013) used cross-sectional data for center-based preschool programs to examine the relationship

between measures of program quality typically included in QRISs and children's developmental outcomes. The study also simulated alternative rating summary scales. Their analysis found that two process-related quality indicators in their generic QRIS—a measure of teacher-child interactions based on the CLASS and a measure of the learning environment based on the ECERS-R—were predictive of children's learning, but the other structural measures were not (e.g., staff quality, ratio and group size, and family partnerships).

Other nonexperimental and quasiexperimental research has found small or no differences in children's outcomes based not only on structural measures of quality but also on process measures (Burchinal, 2010; Burchinal, Kainz, and Cai, 2011; Auger et al., 2014). An illustration of these findings is the evaluation of Boston's high-quality preschool program for four-year-olds, which was found to be effective in raising children's language, literacy, numeracy, and mathematics skills during the preschool year (Weiland and Yoshikawa, 2013). Three common process measures used to rate classroom quality—the ECERS-R, CLASS, and the Early Language and Literacy Classroom Observation (ELLCO)—yielded small or no associations with children's gains in vocabulary and executive functioning (Weiland et al., 2013). The authors concluded that "the current measures of quality may simply not be strong measures of the classroom quality factors that improve children's academic outcomes" (Weiland et al., 2013, p. 207).

Recommendations for OEL and Other Stakeholders

The goal of the RAND evaluation of Delaware Stars was to provide OEL and other stakeholders in Delaware with independent, objective, and rigorous empirical evidence of the extent to which Delaware Stars rating tiers reflect relevant differences in the quality of home- and center-based programs and whether the system is operating well in terms of technical assistance, financial supports, and other features. Such evaluations provide opportunities for continuous improvement of the QRIS itself. Based on our findings—with the limitations of our analysis in mind and the context for the findings in relation to other research on ECE quality and QRISs—we offer several recommendations for OEL as it seeks to implement an effective, robust system for measuring and reporting on the quality of early learning and care programs in home and center settings and for improving quality in ways that are beneficial for participating children and their families.

Learn from Other QRIS Validation Studies

As noted in Chapter 1, more than 20 states have QRIS evaluation efforts under way, although not all are intended to be as comprehensive as the RAND study of Delaware Stars. QRIS validation studies published to date with similar objectives to this one are summarized in Tables 1.5 and 1.6. In the next several years, the set of QRISs with similar evaluation evidence will continue to grow as more of the evaluations supported by RTT-ELC grants are completed. This growing

body of research, much of it using similar measures of child development and program quality, provides an opportunity to look beyond the findings of any one study for any particular QRIS and discern any broader findings with relevance for QRIS design and implementation.

Research syntheses are needed, and perhaps formal meta-analyses as well, to determine why some state rating structures appear to capture quality differences while others do not. This body of research evidence may also shed light on other aspects of system performance, such as the roles that the nature and intensity of quality improvement supports and financial incentives play in advancing classroom practices and program quality.

Consider Further Refinements to the Delaware Stars Rating Structure

If the goal for the Delaware Stars rating system is to capture differences in quality that are important for children's development, the analyses reported in Chapters Three and Four suggest that there is room to refine the rating system beyond the changes that have been phased in since 2014 (e.g., the shift to essential standards). Drawing on the findings reported here and on lessons from similar studies of QRISs, one direction would be to simplify and streamline the rating system with a more limited set of possibly improved measures of the dimensions of quality that matter most for achieving the goals of the QRIS, such as improvements in child developmental outcomes. For example, quality standards related to the management and administration of an ECE program may help to ensure that the ECE program follows sound management practices and may therefore be more likely to stay in business, but those business practices may not have direct implications for children's learning. If the goal of a QRIS is solely to improve children's outcomes, it may be optimal to focus the rating system exclusively on only those factors that are directly related to children's development. Another direction would be to raise the quality standards required to reach the highest rating level, so that programs that reach Star 5 consistently perform at the highest quality levels. Furthermore, consideration should be given to the psychometric properties of the rating scales, using best practices that are emerging in specifying the dimensions of quality to be included in a QRIS rating and how those dimensions are combined to form a single rating (Burchinal, Tarullo, and Zaslow, 2016).

If a global quality measure is to be used (like the ERS in Delaware Stars or the CLASS in some other QRISs), then consideration should be given to using subscales of existing measures that are most strongly predictive of children's development (e.g., Instructional Support in the CLASS). If additional quality standards beyond a global scale are to be used (as is the case with the essential standards and other points-based standards in Delaware Stars), then consideration should focus on the set of standards that are most essential for children's outcomes (or any other goals for the QRIS). Further, attention should be given to how the standards are measured or verified. For example, it may not be sufficient to check a box that a provider is using a developmentally appropriate curriculum. Rather, it may be necessary to verify that the staff are appropriately trained on the curriculum and that it is being implemented effectively on a daily basis. Verification of implementation of high-quality practices with fidelity is likely to be more

challenging and time-consuming than a checklist approach. This is another area in which Delaware can benefit from research in other states and at the national level regarding advances in the measurement and verification of high-quality ECE practices.

Modifying Delaware Stars, even to simplify it, entails costs. The process takes time in terms of designing any modifications, and a pilot may be in order as an initial test of any new rating structure. Obtaining buy-in from providers, parents, and other ECE stakeholders is critical as well. But particularly if the modifications result in a simplified rating scale—such as fewer points-based standards or the use of more-streamlined global quality scales—there are likely to be cost savings that could be redirected toward enhanced quality improvement supports (e.g., TA, coaching) or financial incentives. For example, a simplified rating system would reduce the cost of ratings determination and reduce the burden of ongoing data collection for OEL. The time savings for ECE providers and their staff from a simplified rating system would not show up as explicit dollar benefits, but the opportunity to use that time for purposes that directly benefit children would likely be viewed as a welcome change. A simplified system would also help increase the transparency of ratings for parents. Finally, a simplified rating system could provide a stronger signal to providers about OEL's recommended pathway for quality improvement.

Strengthen Quality Improvement Supports in Delaware Stars

As Delaware Stars has continued to incorporate new center- and home-based providers, there is an ongoing need to support providers in their efforts to improve quality and advance toward higher ratings. The information gleaned from the director interview reported in Chapter 5, while indicating an overall high level of satisfaction with TA supports and financial incentives, identified some areas for improvement. With respect to TA, some directors expressed concern with the consistency in the advice they received as they worked toward a higher rating tier and with the extent of turnover among TA staff assigned to their program. Directors also expressed a desire for more time with their TA provider. All three issues suggest scope for lowering caseloads for TA providers, increasing TA training, and targeting training for various staff who interact with directors to increase the consistency of guidance TA providers offer. In effect, the role for the generalist TA provider could be solidified as a system navigator—an intermediary who connects programs to specialty TA providers, helps notify programs of grants or other financial supports they might qualify for, and is a general expert on the requirements for licensing and for quality improvement in Delaware Stars.

While ECE directors' positive perceptions of TA are an important outcome, the next step in a future evaluation would be to revise the way data is collected regarding TA to allow for analyses of whether the frequency and content of TA relates to improved Delaware Stars ratings or to improved performance in specific areas, such as teacher-child interactions or teaching and learning. Likewise, in the case of financial incentives, given the magnitude of the investment in the various types of incentives, it is important to continue to monitor which providers are accessing those benefits and to investigate the effects they have on improving or sustaining

quality. If some types of incentives prove more effective than others, this may justify shifting resources from one part of the incentive system to another.

Strengthen the Marketing of Delaware Stars to Families

From the evidence assembled in this evaluation, parental recognition and understanding of Delaware Stars appears to be low. Outreach and knowledge building is an important centralized role for OEL. If OEL is going to make further modifications to Delaware Stars, we recommend that marketing should wait until the revised structure is in place. But when the timing is right, our findings suggest that OEL should take stock of current marketing activities and evidence of their impact, compare those strategies to best practice guidance in the field (see, for example, Swanson, 2013; National Center on Child Care Quality Improvement, undated), and determine where new approaches may be called for. This is another area where lessons may be learned from the efforts of other states and localities to market their QRISs.

Enhance Administrative Data Systems to Support Ongoing System Monitoring and Quality Improvement

The two prior RAND reports on the evaluation of Delaware Stars noted the value of administrative data for monitoring the outputs of the QRIS and evaluating its performance. Given the time and expense to undertake the collection of primary data, as was done for this last phase of the RAND evaluation, it is important to exploit existing data generated by the licensing system and by Delaware Stars for assessing performance. Administrative data can be used to track program participation rates, rating levels and changes in ratings over time, and the share of children enrolled in participating programs overall and by rating level. Administrative data can also be used to understand which providers are accessing TA and financial incentives and whether there are meaningful differences in how resources are allocated based on provider characteristics.

Further linkages across data systems can support additional analyses of system effectiveness. For example, data can also be combined to examine whether TA supports, financial incentives, or other aspects of the QRIS influence how rapidly providers move up the rating tiers. Information on which members of the ECE workforce receive financial incentives can be linked with a workforce registry to assess the impact of the incentives on tenure with a particular ECE program and retention in the ECE field more generally. Information on children's participation in ECE programs can be linked to their elementary school records to look at the relationship between the quality of the early learning program and kindergarten readiness (see, for example, the study by Sabol and Pianta, 2012), as well as performance in school at each successive grade. Robust integrated data systems are a key element for facilitating these types of analyses. However, as documented in the first two RAND reports, the Delaware Stars database as currently configured has significant shortcomings. The primary one is that it does not compile some of the most important information that OEL could use to monitor Delaware Stars. These

data include refreshed enrollment data for the entire set of licensed programs, which makes it challenging for the state to regularly compile reports of enrollments by program type and Delaware Stars status. In addition, the database is cumbersome, with hundreds of variables that are routinely missing or are measured in nonstandard ways. Prompted by earlier findings from this evaluation, OEL has been investing in improvements in the Delaware Stars administrative data systems.

To further this objective, we recommend that OEL invest in a dedicated data manager to redesign the Delaware Stars database to reduce substantially the number of data elements collected, while increasing the accuracy and standardization of the data elements that remain. For example, standardizing TA codes and having fewer of them would aid in monitoring the types and amounts of TA provision. Standardizing the tracking of grants and other financial supports to programs is essential for aggregating them into a single database and understanding which programs are receiving various resources and the number and types of children those programs serve. Additional resources may be required for other agencies to enhance the data systems they maintain that feed into the Delaware Stars database, such as the OCCL database of licensed programs.

The ultimate goal is a fully integrated QRIS administrative data system that links a more routinely refreshed state licensing database, along with Head Start and ECAP enrollment data, to the database of providers participating in Delaware Stars. Ideally, the database for Delaware Stars would integrate all information about a provider's rating history, receipt of TA and the types of TA supports, and the allocation of all types of financial incentives to the program and its staff. Provider data would also have at least biannually refreshed information on program enrollment, status with respect to sources of public funding (e.g., Head Start, ECAP, POC subsidies), and accreditation status. A valuable but complex extension to the database would be to collect rosters of children enrolled in each program and assign to each child a unique child ID, thereby allowing the linkage of school-age children back to their ECE programs in the years prior to kindergarten entry. Linkages between ECE providers and an ECE workforce registry would also add complexity but would further expand the analytic value of the data system.

The unifying theme of these recommendations is to use data and other information as part of a strategy of continuous improvement for Delaware Stars. This would include building from this baseline study to further validate the Delaware Stars rating structure, either when the current structure is fully phased in (i.e., when all programs have been validated using the new essential standards) or following any additional modifications to the rating structure. As discussed above, there is also scope to assess the effectiveness of quality improvement supports through TA and other specialized training and to evaluate the impact of the various types of financial incentives.

Recent thinking on QRISs by early childhood experts and practitioners has broadened the QRIS logic model to consider other potential benefits from implementing a QRIS (Zaslow and Tout, 2014; Schilder et al., 2015). Other hypothesized benefits of a QRIS include professionalization of the ECE workforce, the development of a strong ECE system,

enhancement of parenting and other aspects of family functioning, and promotion of financial viability and stability of ECE providers. If there are broader goals, then future research could determine whether the rating scale is predictive of these other goals, such as a more professionalized workforce (e.g., as measured by reduced turnover), improved parenting practices, or reduced rates of ECE program closures.

A focus on continuous improvement for Delaware Stars comports with the growing recognition of the importance of using data, analytics, evidence, and evaluation—known as the "moneyball" approach—to provide regular feedback mechanisms for assessing the current landscape, identifying where improvements are needed, implementing the needed modifications, and then monitoring and evaluating further (Mead and Mitchell, 2016). This amounts to building a culture of learning and improvement that permeates all levels of the early learning system, from the micro level (e.g., teachers, classrooms, and providers) to the broader ECE system level (e.g., child care licensing, QRIS, and professional development systems). With the set of action steps represented by these recommendations, OEL can model such a data-driven approach to investing in an ECE system that provides the desired benefits for children and their families, as well as ECE providers and the ECE workforce.

Appendix A. Additional Documentation for Delaware Stars

This appendix provides additional information about the structure of financial incentives for Delaware Stars as of July 2014.

Table A.1. Financial Incentives Available to Delaware Stars Providers or Staff as of July 1, 2014

Financial Incentive	Description	Features
Provider-Based Incentives		
QI Grants	Provides grants to programs at Star 2, Star 3, or Star 4 that are tied to the provider's QIP; funds can be used for professional development, materials, or other projects; programs can receive this grant once per star level	Small FCC ..... up to $750 Large FCC ..... up to $1,000 Small center (13–60) ..... up to $2,500 Medium center (61–120)..... up to $3,000 Large center (121–200) ..... up to $4,000 Extra-large center (201–300)..... up to $5,000
Infrastructure Fund	Provides grants to programs at Star 2 or higher for capital or technology improvements that will support moving to the next star level; programs may apply for either or both awards and may receive more than one award over time	Requirements: Improvements must be related to a program's QIP; programs must participate in POC and serve high-need children
POC Tiered Reimbursement Bonuses	Provides an escalating reimbursement rate for POC subsidies for providers at Star 3 or higher, with the following tiered structure for children up to age five: Star 3: 80 percent of market rate Star 4: 93 percent of market rate Star 5: 102 percent of market rate	Requirements: Valid contract with Division of Social Services for POC
Staff-Based Incentives		
CORE awards	Provides grants to early childhood educators employed in Star 3, Star 4, or Star 5 programs for making gains on the Delaware Early Childhood Career Lattice (awards from $500 to $6,000), for being recruited to the field ($1,000), and for remaining in the field ($2,000 to $3,100)	Eligibility for degree/credential awards: Administrators, teachers, assistant teachers, curriculum coordinators, and FCC providers who have reached Step 4 or higher on the Career Lattice and are employed for at least 30 hours per week in Star 3 to Star 5 programs and meet a specified wage threshold Eligibility for recruitment awards: Newly recruited teachers who have reached Step 7 or higher on the Career Lattice and are employed for at least six months in a Star 3 to Star 5 large FCC or center and meet a specified wage threshold Eligibility for retention awards: Administrators, teachers, assistant teachers, curriculum coordinators, and FCC providers who have reached Step 8 (administrators) or Step 7 (all others) or higher on the Career Lattice and who are employed for at least 12 months at the same Star 3 to Star 5 program and meet a specified wage threshold

Table A.1. Financial Incentives Available to Delaware Stars Providers or Staff as of July 1, 2014, Continued

Financial Incentive	Description	Features
T.E.A.C.H. Early Childhood®	Provides scholarships and support to ECE professionals pursuing a degree or credential related to early childhood education at a participating Delaware higher education institution	Eligibility: Individuals enrolled in programs that grant degrees in early childhood or child development

SOURCES: Delaware Institute for Excellence in Early Childhood, undated; and OEL, 2014.
NOTE: Merit awards are excluded because they were no longer available as of July 1, 2014.

Appendix B. Data Collection Procedures and Response Rates

This appendix documents the procedures used to collect data for the provider sample and the child sample. This includes the process for recruiting ECE programs to participate in the study; for recruiting, training, and managing field staff; for conducting child developmental assessments; and for conducting observations of program and classroom quality. Response rates for the provider sample are also provided, along with documentation of the construction of analytic weights to account for nonresponse.

Recruitment of ECE Programs into the Provider Sample

RAND obtained a list of all licensed ECE providers in Delaware as of May 2014, combined with the 25 school-based preschool programs in Delaware Stars as of that date. After excluding providers identified as serving only school-age children, the list had 1,196 providers. As shown in Table B.1, this included 454 providers participating in Delaware Stars and 742 providers not participating in Delaware Stars. Providers were stratified into four groups based on their type of license: licensed centers, licensed large FCCs, licensed small FCCs, and license-exempt school-based preschool programs. The total number in each category as of May 2014 is shown in Table B.1.

With a goal of collecting information from 200 to 300 providers in total and with an expected response rate of 60 percent, we determined that we needed to invite 100 percent of all providers in Delaware Stars to participate in the study. With the exception of small FCCs, we also opted to invite 100 percent of centers and large FCCs not in Delaware Stars. Given the large number of small FCCs not participating in Delaware Stars (596 providers), we randomly sampled 29 percent of those providers, or 174 programs. In total, 774 providers were invited to participate in the study, or about 65 percent of the potentially eligible providers.

In May 2014, the RAND Survey Research Group mailed via the U.S. Postal Service an initial recruitment packet to the sampled providers. Our initial recruitment packet included an invitation letter, a consent form, a brochure about the study, and an offer of a $100 gift card for participation. RAND field staff then placed several follow-up calls to the programs throughout the summer of 2014. However, with lower response rates than anticipated, RAND increased the participation incentive from $100 to $200 in August 2014 and mailed (via FedEx) another invitation packet to all nonresponding programs. From August to December 2014, field staff then placed in-person visits as well as repeat phone calls to nonrespondents. Some sampled ECE programs were part of a school district or centrally managed as part of a multi-site agency. These cases sometimes required approval by a higher authority (e.g., school district board or agency manager) prior to the individual program being able to give consent to participate. In many of

these cases, the project director placed a personal call, and field staff followed up with an in-person visit to collect consent forms. As detailed in the concluding section of this appendix, 181 programs participated in one or more of the data collection components, for an overall response rate of 25 percent among invited and eligible providers. The construction of analytic weights to account for nonresponse is discussed in Appendix D.

Table B.1. Sample Frame and Sampling Rates for the Provider Sample

Provider Type	Number as of May 2014	Number Invited to Participate in Study	Sampling Rate (Percentage)
Total	1,196	774	64.7
In Delaware Stars	454	454	100.0
Licensed centers	241	241	100.0
Licensed large FCCs	26	26	100.0
Licensed small FCCs	162	162	100.0
Public schools	25	25	100.0
Not in Delaware Stars	742	320	43.1
Licensed centers	99	99	100.0
Licensed large FCCs	47	47	100.0
Licensed small FCCs	596	174[a]	29.2

SOURCE: Authors' disposition file.
[a] Randomly selected.

Recruiting, Training, and Management of Field Staff

As detailed in Table 2.2, data collection took place in three phases:

- **Fall 2014:** director interview and child developmental assessments
- **Winter 2015:** observations of quality in small and large FCCs and school- and center-based ECE classrooms
- **Spring 2015:** second wave for director interview and child developmental assessments.

Data collection was preceded by field staff recruitment in the fall of 2014, with three training periods corresponding to each wave of data collection.

Field Staff Recruitment

Field staff were recruited with the assistance of a staffing services company with which RAND routinely works for field data collection projects. A job description was provided by RAND to the agency. The criteria for the field staff included experience with children; ability to pass a Megan's Law background check; a valid driver's license; good communication and organization skills; and previous experience in survey administration, customer service, or research. In addition, a preference for bilingual fluency in Spanish and English was included in the job

posting. The agency performed an initial screening of potential candidates and then forwarded the candidates to RAND for further review by RAND Survey Research Group staff. To reach ECE programs throughout the state, staff were specifically recruited from different geographic areas around the state of Delaware. A total of 16 individuals were recruited for the fall data collection process. Field staff who applied as bilingual candidates were assessed for verbal and written Spanish proficiency. Field staff who passed all rounds of interviews and background checks were invited to training.

Fall 2014 Training for Child Assessments

A total of 16 field staff were invited to training for the fall data collection. The training was conducted over a six-day period. Staff were sent training binders to review prior to the first day of training. The binders contained information concerning the training agenda, project overview, confidentiality and scientific misconduct standards, general interviewer training, project protocols, and child assessment and study materials.

The first training session was a telephone and webinar conference, which gave an overview of the study, discussed research ethics (data safeguarding, informed consent, confidentiality, scientific misconduct), provided general interviewing techniques, explained key study information, and covered general data collection processes. Phone and webinar training was followed by four days of training at a hotel conference facility in Delaware, where the field staff were given detailed instructions on study protocols, how to administer and score the child assessments, and how to complete such administrative tasks as the child roster and child information sheet. Field staff were provided with wheeled boxes containing all materials that they would need for conducting data collection activities at each site, including assessment manuals, response booklets, and "thank you" books for participating children.

In-person training sessions focused on the correct administration of the five child assessment tools, a step-by-step plan for conducting site visits, and administrative activities around the data collection. In order to ensure that the field staff were completing the assessment appropriately, they were also given ample time to practice assessment administration with each other. Throughout the practice administrations, the field staff were monitored by and received feedback from RAND study staff. At the end of these four days, two of the 16 recruited field staff were dismissed from the project, as they did not meet the requirements for collecting reliable data.

Immediately following the training sessions, the field staff were given an opportunity to practice administering child assessments at a large center enrolled in the study. During the practice sessions, RAND study staff guided the field staff on what to do when they arrived at a program and provided feedback on administration of the child assessments. Field staff members administered child assessments to eligible children while fellow field staff and RAND study staff watched from an observation room. RAND study staff certified that field staff were ready for data collection. At the end of the practice sessions one additional field staff member resigned, leaving 13 staff to begin the fall data collection.

Additional training for Spanish language assessments was conducted by bilingual RAND study staff. The training session covered the three Spanish assessments instruments: TVIP, WJ Batería, and HTKS–Spanish. Bilingual field staff were monitored during several practice sessions.

A separate training session was held for field staff who were responsible for administering the director interviews. The field staff were instructed by a RAND study staff member on general interviewing principles and technical skills, including how to log on to the computer, access the RAND network, and retrieve and return interview cases using the CASES software. During this training session, the field staff practiced the director interview with one another. Field staff were required to pass a mock interview with a RAND study staff member prior to beginning director interviews.

Winter 2015 Training for Observations of Program Quality

Eleven of the original field staff remained with the study for the next round of training; four new staff were recruited, passed background checks, and were invited to the training for ECE program observation. The training took place over 12 days during January and February 2015. Training was in person and took place in a central location in Delaware. Training involved a combination of vendor-led sessions and RAND study staff review and clarification meetings.

Teachstone training personnel conducted the CLASS Toddler and CLASS Pre-K trainings. The training for each CLASS tool lasted two days. Field staff attended training and passed Teachstone's online reliability test for each CLASS tool prior to conducting observations. For each reliability test, field staff watched and coded five 20-minute online videos.[32] Field staff had to obtain a minimum of 80-percent agreement within one point (on the seven-point scale) of the master codes across all five videos, and at least two codes (out of five) had to be within one point of the master code for each dimension to meet reliability standards for each CLASS tool. Teachstone allows up to three attempts to pass the reliability test; staff failing to meet the reliability standards after three attempts did not conduct CLASS observations for this study. RAND project staff and Teachstone personnel were available for assistance and review during the certification process.

HighScope personnel conducted training for the PQA Infant/Toddler, Preschool, and FCC observation tools. Training occurred for five days. Like the CLASS, field staff had to attend training and pass HighScope's reliability test for each PQA tool prior to conducting observations. Each PQA reliability test included video clips, written scenarios, and scoring decision questions.[33] Staff had to achieve a score of 80 percent correct on each section of the test within two attempts to pass the test for the specified PQA tool.

---

[32] The reliability test, video clips, and master codes are part of the CLASS training materials prepared by Teachstone.

[33] The reliability test, video clips, and master codes are part of the PQA training materials prepared by HighScope.

Following the vendor training sessions, RAND project staff reconvened the field staff to provide additional training on the measures, conduct training for the CIS, review administrative and rating forms, and provide a step-by-step procedure for the observations. This additional RAND training lasted three days.

Spring 2015 Training for Second Wave of Child Assessments

For the second round of child assessments, experienced field staff from the fall data collection received daylong refresher training in which testing procedures were reviewed and practice sessions were conducted. Field staff who did not conduct assessments during the fall data collection received the full round of training on the assessments and data collection procedures and were accompanied by RAND study staff on their first few provider visits.

Management of Field Staff

The fall 2014 data collection period started in October 2014 and ended in February 2015. After each site visit or completed interview, the field staff emailed a report to the RAND study staff documenting the date the center was visited, the number of permission forms received, the number of children rostered and assessed, forms collected, status of the director interview, and the scheduled return date if needed. Weekly management calls were held with field staff members to provide updates on project progress, discuss the upcoming schedule, review updates or changes to the data collection plan, and discuss difficulties or problems with the fieldwork.

RAND study staff observed the field staff during their visits to the ECE programs to ensure that the assessment instruments were administered correctly. Staff completed an evaluation form documenting whether field staff followed the appropriate procedures, the number of sessions observed, and notes on the observed session. Monitoring occurred at approximately 15 percent of providers, each field staff member was monitored on multiple occasions, and assessments for over 7 percent of the children were observed. The RAND project staff provided feedback and refresher training as necessary to the field staff.

The winter 2015 data collection period started in February and ended in July. As in the fall, after each scheduled visit to a provider site, the field staff sent an email report to RAND study staff. The report detailed the type of classrooms and the number of classrooms observed. In addition, the field staff confirmed whether any of the classrooms did not exist. Weekly telephone conferences were held with the staff to provide any updates and to discuss concerns and questions about the observations.

For the observations, RAND study staff accompanied field staff to observation sites, remained in the classrooms for a portion of the observation time, and reviewed the field staff's notes and scoring after the observation was completed. In addition, the RAND study staff documented the field staff member's observation techniques and patterns. RAND study staff visited approximately 8 percent of the classrooms that were observed by the field staff. Clarification and retraining was provided as needed.

In addition to the monitoring by the RAND study staff, field staff were required to complete a recertification of the CLASS and PQA measures on which they were certified to address drift and recalibrate their scoring. The recertification occurred after the field staff completed approximately ten observations using each specific measure or every three weeks (whichever came later). Staff had up to two attempts to pass the online certification tests for each CLASS and PQA tool.[34] Each test was a shortened version of the reliability tests described previously—i.e., each CLASS calibration test required coding one 20-minute video and each PQA calibration test included fewer written scenarios and scoring decision questions (no videos). To pass calibration tests, staff had to meet the same reliability standards described previously for the CLASS and PQA.

The recertification test was generated and scored by either Teachstone or HighScope. The field staff were considered recertified if they scored 80 percent or higher on the test. The field staff were able to observe using the measure that required recertification only after they passed the test.

The spring data collection period began in April 2015 and ended in August 2015. As in the fall, after each site visit the field staff prepared and emailed to RAND a report documenting the status of the data collection. Weekly management calls continued through spring data collection. RAND study staff observed the field staff during their visits to the provider sites to ensure that the child assessment instruments were administered correctly. Monitoring occurred at approximately 12 percent of providers, each field staff member was monitored on multiple occasions, and assessments for over 5 percent of the children were observed.

Procedures for Rostering Children and Conducting Developmental Assessments

In fall 2014, RAND mailed a set of parent informed-consent packets to the center and large FCC providers that had consented to be in the study. Each packet consisted of an invitation letter to the parent, a consent form describing the child assessments to be conducted in fall 2014 and spring 2015 and other information required for informed consent, a study brochure with questions and answers, and a manila envelope in which to return the signed consent form. The consent form also included a short survey to collect essential child and family background information (see Appendix C for the consent form and parent survey). Program staff were instructed to distribute these forms to parents of all of their enrolled children in the target age range (birth dates from September 1, 2009, to August 31, 2012, approximately ages two to five). The center director was asked to not open any of the envelopes returned by parents, but rather to hold the envelopes for pickup by RAND study staff.

[34] Teachstone prepared the CLASS calibration tests, and HighScope prepared the PQA calibration tests.

To minimize the burden on providers, the data collection protocol and provider consent specified assessments for up to 15 children per site. On the day of the first site visit, the field staff opened and reviewed each returned parent consent form to verify that the child was eligible for assessment. The criteria were as follows: The parent had signed the consent form; the child's birth date was in the range from September 1, 2009, to August 31, 2012; and the child was currently enrolled in the program. Next, if the child was eligible, the consent form was reviewed for completeness. Consent forms with a completed parent survey were prioritized. As a last step, children were added to a roster in the order that the consent forms were received, with children listed into one of the three kindergarten entry cohorts based on their birth date: oldest cohort (born September 1, 2009, to August 31, 2010), middle cohort (born September 1, 2010, to August 31, 2011), or youngest cohort (born September 1, 2011, to August 31, 2012). The roster listed the child's name, date of birth, language spoken at home, and teacher/classroom. Final assessment outcome, outcome date, and any case-specific notes were annotated for each child rostered. If there were more than 15 children with parental consent, children were added to the roster such that they would be equalized across three age cohorts and, within a given age group, be in the same classroom where possible.

A total of 2,026 parent permission forms were received, but because not all were for age-eligible children, not all consent forms were signed, and some programs had more than 15 consents, a total of 1,519 children were ultimately rostered.

The field staff also had a form (Information About Children Form), to be completed by the director or his/her designee, that collected enrollment counts for the program and information about the rostered children. Prior to providing the form to the director, the field staff filled out the "Name of Child" section. The form had fields to record the following information for each rostered child: gender, date of birth, month and year child first enrolled in the program, receipt of POC, participation in ECAP, hours per week and days per week child attended the program, scheduled days absent in past month, child's classroom name or number, days and times child attended the program (for scheduling the assessments), whether the child had a diagnosed disability, and language spoken at home. The form was provided to the director with an envelope, and the field staff requested that it be ready to be picked up at the end of the day or during their follow-up visit, if needed.

The field staff coordinated with the provider staff and director to bring rostered children to a quiet space set aside for the assessments.[35] Before the child was assessed, he or she was read an assent statement by the field staff and asked for verbal assent. If the child did not provide assent, he or she was escorted back to the classroom. For such children, a second attempt was made later the same day or on another day. If the child refused again, a replacement child was added to the

[35] In some cases, the program staff indicated to the field staff that a child was developmentally disabled and could not be assessed. If they indicated this, the field staff selected a replacement child from the group of children whose parents had signed the consent form and who had not already been rostered. Priority was given to a child in the same age cohort and classroom as the child with a diagnosed disability.

roster, if one was available, from the group of children whose parents provided consent but had not been previously rostered. Priority was given to a child in the same age cohort and classroom. If the child was initially absent, the field staff returned the site at least two times to assess the absent child. Again, if the child could not be assessed, a replacement child was added to the roster using the same protocol as cases where assent was not given.

Field staff then administered the set of four direct or performance-based assessments one on one: PPVT, WJ–Applied Problems, WJ–Letter Word Identification, and HTKS.[36] The order of the assessments was randomized using color-coded packets for each sequence (for example, in green-colored packets, the PPVT came first, followed by WJ–Applied Problems, WJ–Letter Word Identification, and HTKS).[37] (The color of the row on the roster where the child's name happened to appear determined the color of the packet used by the field staff.) At the close of the assessment session, children were given a book in appreciation for their participation and returned to their classroom. Books were available in English and Spanish. In addition to the direct assessments, the DECA instrument for each child was provided to his or her teacher or caregiver, to be completed by the end of the day.

The conduct of the child assessment in the second wave followed a similar protocol. All children who were assessed in the first wave were eligible for a second wave of assessment. The schedule of provider visits was established, to the extent possible, to maximize the time between the fall and spring assessment. Again, children were asked to provide verbal assent before the assessment began, with a second attempt made for those who did not provide assent on the first attempt. The assessments were administered in the same order as the first wave (in which the order was randomly determined). Up to two visits were made to a provider site in order to complete the assessments for all eligible children. These procedures for the fall and spring assessment waves resulted in 1,315 children assessed in the first wave and 1,123 children reassessed in the second wave.

Procedures for Provider Observations

Between February 2015 and July 2015, RAND field staff observed small and large FCCs and school- and center-based programs to assess program quality. For small FCCs, the observation was for the home-based setting using the age-appropriate instrument as detailed in Appendix C. For large FCCs and centers with more than one classroom, the provider consent allowed for observation of up to three classrooms per site. RAND sought to observe up to one toddler classroom and up to two preschool classrooms. In programs where there was more than one toddler classroom or more than one preschool-age classroom, RAND sought to observe those

[36] The procedure for assessing children in Spanish is described in Appendix C.

[37] The HTKS was always administered last because of the need to have a good rapport with the child in order to have a successful administration. In total, there were six different orderings of the other three assessments.

classrooms where the greatest number of children were enrolled who had been assessed in the fall. In particular, in advance of the site visit, a classroom observation roster was completed, listing the classrooms to be observed in order of preference.

Upon arrival, the field staff confirmed with the director the classrooms that would be observed during the visit. When entering a classroom, the field staff read a consent statement to the teacher and obtained oral consent before starting the observation. When the first-choice classroom could not be observed (because the director or teacher refused, the teacher was no longer present, or the original classroom had been dissolved), the field staff would move down the list to choose an alternate classroom. Overall, 93 percent of the classrooms observed in centers were the prioritized classroom on the roster.

As detailed in Appendix C, the CLASS and PQA require the use of different instruments, depending on whether the site is home- or center-based and depending on the ages of the children in the group. During a visit, one observer completed the PQA, while the other collected the CLASS, the CIS, and the counts for measuring group size and staff-child ratio. Field staff who were certified on both the PQA and CLASS could complete all measures in a single visit. Each observer completed an End-of-Visit form at the conclusion of the observation. In some cases, the observation was conducted over two days if schedules could not be coordinated to have two observers certified on the different measures on the same day. The most common hours for observation were from 8 a.m. to 1 p.m., but exact hours varied depending on the program's schedule.

Provider Response Rates

Table B.2 summarizes the response rate calculations for the provider sample, in total and separately for providers in Delaware Stars and not in Delaware Stars, with additional breakdown by provider type. As noted earlier, 1,196 providers were potentially eligible for the study. After random sampling of 29 percent of small FCCs, 774 providers were invited to take part of the study. A total of 181 programs completed at least one data collection component, for an overall response rate of nearly 25 percent. The refusal rate was 38 percent. It is important to note that participation in the validation study was considerably higher for providers in Delaware Stars than for those not in Delaware Stars (36 percent versus 8 percent; results not shown). The response rates were highest and nearly equal for centers and public school programs in Delaware Stars (46 percent and 48 percent, respectively). The response rate for centers not in Delaware Stars was about half the rate as that achieved for centers in Delaware Stars (23 percent versus 46 percent).

Of the 181 providers in the final study sample, all completed at least one director interview. Excluding the small FCCs, which were not eligible for the child assessments, there are 143 providers among the 144 large FCCs and school- and center-based programs with child assessments. All but nine of the 181 providers participated in the observations of program quality.

Table B.2. Response Rates for the Provider Sample

Measure	Providers in Delaware Stars				Providers Not in Delaware Stars			Total
	Small FCCs	Large FCCs	Centers	Public Schools	Small FCCs	Large FCCs	Centers	
Licensed programs as of May 2014	162	26	241	25	596	47	99	1,196
Programs invited to participate in study[a]	162	26	241	25	174	47	99	774
Not eligible	10	0	11	0	12	5	8	46
Duplicate listing	0	0	1	0	0	0	0	1
In pilot sample	1	0	1	0	0	0	1	3
Out of sample; no age-eligible children	9	0	9	0	12	5	7	42
Eligible	152	26	230	25	162	42	91	728
No data collection	118	22	125	13	160	41	71	550
Refusal	50	11	87	9	64	13	41	275
No contact (consent or refusal)	67	11	37	4	96	28	29	272
Data collection completed	35	4	106	12	2	1	21	181
Fall and/or spring director interview	35	4	106	12	2	1	21	181
Fall and/or spring child assessments	–	4	106	12	–	1	20	143
Winter quality observation	33	4	103	12	2	1	17	172
Refusal rate (%)	32.9	42.3	37.8	36.0	39.5	31.0	45.1	37.8
Response rate (%)	23.0	15.4	46.1	48.0	1.2	2.4	23.1	24.9

SOURCE: Authors' disposition file.

NOTES: Response rates calculated per methods set forth in American Association for Public Opinion Research, 2015. – = not eligible.

[a] Selection was random when fewer than 100 percent of providers were invited to participate.

Construction of Analytic Weights

We developed and applied weights for the final provider sample to make inferences that generalize to all licensed providers in Delaware. The weights account for the sampling of providers for the initial invitation (e.g., the sampling applied to small FCCs), nonparticipation at the time of recruitment, and attrition from those providers that originally provided consent but then chose not to complete data collection.

To develop the weights, we augmented the August 2015 frame of licensed ECE providers with an indicator for those providers that participated in data collection. The database provided information on provider type and participation status in Delaware Stars (including the rating for those in Delaware Stars). In addition, using the provider's zip code, we matched each provider to U.S. Census data measuring the population distribution by race-ethnicity, the distribution by education level, the rate of female-headed families, median household income, and the poverty rate for children under age six.

With these key characteristics, we constructed sampling (including nonresponse) weights. Specifically, we estimated logistic regression models to predict the propensity of a provider in Delaware participating in the study as a function of provider type, Delaware Stars status as of August 2015, provider county, and the set of provider zip code–level variables. Provider type and Delaware Stars status were significant predictors of which providers participated in the study. For example, response rates were higher for centers and school-based programs compared with small and large FCCs. Providers in Delaware Stars were also more likely to participate in the study compared with those not in Delaware Stars. Among providers in Delaware Stars, response rates were highest for Star 3 and Star 4 programs. The provider's county and zip code–level characteristics were not statistically significant predictors of participation in the study.

We used the inverse of the predicted propensity as sampling weights. Because logistic regression is known to produce extreme values in some cases, we reviewed the distribution of the estimates and trimmed extremely large weights to avoid overly influential observations. The weights took into account the two-stage process of a provider being invited to participate in the study, followed by the invited provider choosing whether to participate. These weights are used as sampling weights in all analyses at the provider level, including those reported in Chapters 2, 3, and 5.

As shown in Table B.3, the weighted distribution of the provider sample on several key characteristics that we can measure for all providers closely matches the distribution for the original sampling frame, with the exception of Delaware Stars status and ratings. Notably, the large number of small FCCs among licensed programs not in Delaware Stars (596 out of 742 providers not in Delaware Stars), combined with our low sampling rate for small FCCs and low response rate, means that we have just 16 small FCCs not in Delaware Stars in our final sample to represent nearly 600 providers. Because we have trimmed the sampling weights, the weighted distribution of providers shows just 19 percent of providers not in Delaware Stars versus the 60

percent rate in the frame. When we examine the percentage distribution across rating tiers for those programs in Delaware Stars, the (conditional) weighted distribution comes closer to the distribution among all providers in Delaware Stars (Table B.3).

Table B.3. Characteristics of Weighted Provider Sample Compared with Provider Frame

Characteristic	Provider Sample			All Licensed Providers
	Number	Unweighted Percentage	Weighted Percentage	
Provider type (percentage distribution)				
Small FCC	37	20.4	34.7	28.4
Large FCC	5	2.8	4.0	6.1
Center	127	70.0	58.6	63.4
School-based	12	6.6	2.8	2.1
Delaware Stars rating as of August 2015 (percentage distribution)				
Not in Stars	16	8.8	34.0	60.2
Starting with Stars	2	1.1	2.3	2.9
Star 2	32	17.7	18.2	11.5
Star 3	22	12.2	10.4	5.5
Star 4	62	34.3	19.2	11.2
Star 5	47	26.0	15.8	8.6
Among providers in Delaware Stars, rating as of August 2015 (percentage distribution)				
Starting with Stars	2	1.2	3.5	7.3
Star 2	32	19.4	27.7	29.0
Star 3	22	13.4	15.8	13.9
Star 4	62	37.6	29.2	28.2
Star 5	47	28.5	23.9	21.7
Provider zip code characteristics				
Race-ethnicity of population (average percentage distribution)				
Non-Hispanic white	–	59.3	57.3	59.0
Non-Hispanic black	–	25.6	28.0	26.5
Hispanic	–	10.0	9.3	9.1
Non-Hispanic other	–	5.1	5.4	5.4
Education of adults age 15 and above (average percentage distribution)				
Less than high school	–	14.4	13.4	14.0
High school diploma or GED	–	33.6	32.5	33.3
Some college	–	26.8	27.2	27.4
Bachelor's degree or higher	–	25.1	26.8	25.3
Average percentage of family households with children under age 18 and headed by single parent	–	18.1	18.6	17.6
Average percentage of children under age six living in families with income below the federal poverty line	–	13.9	13.4	13.4
Average median household income (thousands of dollars)	–	58.1	60.2	58.9
N	181	181	181	1,196

SOURCE: Authors' tabulations of licensed database and provider sample disposition file, matched to U.S. Census zip code characteristics. State administrative data are the source for Delaware Stars participation and license type.
NOTES: Percentage distributions are calculated excluding missing cases. Percentage distributions may not sum to 100 because of rounding.

Appendix C. Data Collection Measures and Instruments

In this appendix, we provide additional detail on the collection of child developmental assessments and measures of ECE program quality. We also include the instruments for the director interview and the parent survey collected as part of the consent process.

Child Developmental Assessments

The direct and performance-based assessments of children listed in Table 2.3 could be performed in English only or English and Spanish following protocols described in this section.

Assessment of Children in English

All children, regardless of age, were assessed using the PPVT–4 (Form A). The first part of the assessment was a practice section: Children had to get two training items correct to move on to the main assessment. Children two years up to three years and 11 months of age were administered training page A, and children four or five years old were administered training page B. If the child answered one of the first two questions incorrectly on page B, the administration moved to training page A, and the child had to get at least two questions correct to proceed to the main assessment. If a child did not get two answers correct on the practice page, the child was classified as "untestable" for the PPVT.

Children two years up to three years and 11 months of age who successfully completed the practice session continued on to Set 1 of the main assessment. Children who were four years old began with Set 2; children who were five years old began with Set 4. Children who started on Set 2 or 4 had to receive no more than one error in order to establish a basal set. In these cases, if two or more errors occurred in the starting set, administration proceeded backward by set until the basal set (a set in which there was at most one error) was established. The test was administered until the child received eight or more items wrong in a full set, therefore establishing the ceiling set.

All children, regardless of age and language, were administered the WJ–III Letter Word Identification (Form C). All children started on item one; correct responses were scored a "1" on the response form, and incorrect responses were scored "0." Field staff read all the words in blue, stayed on script, and did not repeat any of the instructions again to the child. If the child's response was unclear for a question, field staff allowed the child to complete the entire page first and then asked all question items on the same page and scored only the unclear item. For fall administration, field staff ended the assessment when the child received six errors in a row. Field staff then counted the number of correct items and filled out the specified box.

The WJ–III Applied Problems (Form C) was completed by all children, with the exception of those who did not pass the language screener (discussed in the next section). The Applied Problems assessment was administered similarly to the Letter Word Identification with the exception that, unlike the latter, field staff could repeat questions when administering the Applied Problems assessment if the child asked.

The English version of the HTKS was not administered to children two years and 11 months and under, nor to Spanish-language children who did not pass the language screener (discussed in the next section). The HTKS is divided into three parts; each part begins with an introduction and a practice session. During the practice session, field staff corrected the child up to three times (parts 1 and 2) or up to two times (part 3) if he or she did not perform the proper action. Once the practice was completed, the main section was administered. Children followed instructions given by the field staff and were scored 0 if the child did not perform the correct action; 1 if he or she made any discernable motion toward an incorrect response, but self-corrected; and 2 if the child performed the correct action. If the child scored 4 or more on a section, the next section was administered. Scores were added across the sections to obtain the full assessment score.

Assessment of Children in Spanish

If the parent checked on the parent survey that the child's primary language at home was a language other than English and the child was three years or older, he or she was administered the first two subtests of the Preschool Language Assessment Scale (preLAS) (Duncan and DeAvila, 2000) as an English-language screener (Vogel et al., 2008; Rainelli et al., 2014). If the child was under three years of age and the home language was noted as Spanish, the Spanish-language assessments were administered without the initial preLAS screener.

Two sections of the preLAS were used for the study: Simon Says and Art Show. The field staff read the script provided on the assessment form, and if the child answered an item correctly, the staff member placed a check mark in the circle beside in the question. The number of correct answers was marked at the bottom of the section, and then the two numbers were added to find the final preLAS score. Passing scores were 11 and over for children three to three years and 11 months and 15 and over for children four to five years and 11 months. Children with a passing score were considered English-proficient, and only the English-language assessments were administered in fall 2014 and spring 2015. For children who did not pass the preLAS and whose home language was Spanish, the PPVT and WJ–Letter Word Identification were administered in English prior to completing the assessments in Spanish (as described later in this section). If the parent noted that the child's primary language was a language other than English or Spanish, the English assessments were administered in the fall and spring regardless of the score on the preLAS.

Assessments in Spanish were conducted in fall 2014 for children who did not pass the preLAS and whose primary language at home, as reported by the parent, was Spanish. This

included the Spanish-language equivalent of each of the direct or performance-based assessments listed in Table 2.3: Test de Vocabulario en Imagenes Peabody (TVIP; Dunn et al., 1986), the Letter Word Identification and Applied Problems subtests of the WJ Batería III (Woodcock et al., 2007), and a Spanish translation of the HTKS (Cabeza-Dedos del pie-Rodillas-Hombros–Extendido; Cameron and McClelland, 2011, supplied by the authors). (The self-administered DECA was completed by the caregiver or teacher in English.) In spring 2015, these children were assessed in English (PPVT and WJ–Letter Word Identification) and Spanish (all direct and performance-based assessments). The preLAS was administered again but not used as a screener.

The rules and instructions for administration of the Spanish-language versions of the WJ subscales and the HTKS were the same as those for the English-language version. The administration and rules for TVIP were different from those for PPVT. Training included Pages A, B, and C, and children were required to get three consecutive correct responses in order to proceed to the assessment. Children four years and 11 months and younger began on the first item, and children five years and older began on item 10. For children starting on item 10, if one of the first eight items was incorrect, the assessment was administered backward from the starting item until a base (eight consecutive items correct) was established. Correct responses were marked by circling an item number; incorrect responses were marked by drawing a slash through the item number. The assessment continued until the child had six or more errors in the lowest eight consecutive items. Among the 1,123 children tested in both fall 2014 and spring 2015, 55 (4.9 percent) took the Spanish-language Batería Applied Problems, 58 (5.2 percent) took the Spanish-language Batería Letter Word Identification, 36 (3.2 percent) took the TVIP, and 53 (4.7 percent) took the Spanish-language HTKS.

Measures of ECE Program Quality

To assess quality not already included in the Delaware Stars rating system, we conducted the PQA, CLASS, and CIS instruments in up to three classrooms for all provider sample programs. Table C.1 summarizes the quality measures administered in centers and FCCs according to the ages of children present.

For the PQA, we used the version appropriate for the setting observed—e.g., when observing classrooms within centers, we used the Infant-Toddler or Preschool PQA as applicable, and we used the Family Child Care PQA to observe FCC settings. Each version of the PQA includes items relating to the classroom physical environment and interactions—e.g., general sections measure the learning environment (physical objects/equipment and facilitation of instruction), schedules and routines, and adult-child interactions. The center-based PQA instruments (Infant-Toddler and Preschool) include an additional section to assess curriculum planning and child observation practices.

Table C.1. Observation Instruments Collected by Program Type and Child Ages

Observation Instrument	Center Classrooms			Small or Large FCCs		
	Toddlers Only (2:0–2:11 years)	Preschoolers Only (3–5 years)	Mix of Toddlers and Preschoolers	Toddlers Only (2:0–2:11 years)	Preschoolers Only (3–5 years)	Mix of Toddlers and Preschoolers
PQA Form A—Center, Infant and Toddler	✓	–	✓ (When majority were under 3 years of age)	–	–	–
PQA Form A—Center, Preschool	–	✓	✓ (When majority were ages 3 years or older)	–	–	–
PQA Form A—FCC	–	–	–	✓	✓	✓
PQA Form B—Agency Items	✓	✓	✓			
CLASS Toddler	✓	–	✓ (Alternate observation cycles)	✓	–	✓ (Alternate observation cycles)
CLASS Preschool	–	✓	✓ (Alternate observation cycles)	–	✓	✓ (Alternate observation cycles)
Arnett CIS	✓	✓	✓	✓	✓	✓
Group Ratio Coding Sheet[a]	✓	✓	✓	✓	✓	✓
End of Visit Form[a]	✓	✓	✓	✓	✓	✓

NOTES: ✓ = collected. – = not collected or not applicable.
[a] Developed by study team.

Observers spent four to six hours in a classroom to complete the PQA for full-day programs and up to four hours for half-day programs. The PQA tool is organized by topic, and the protocol is fluid—i.e., it is not organized by specific "observation cycles," and all components of the day should be considered to rate an item when appropriate (e.g., assessment of materials available to children can be completed throughout the observation). To streamline the observation and ensure that we observed each transition and activity, we used the classroom schedule to guide our observations. Thus, field staff generally observed all main activities, including arrivals, mealtimes, outdoor play, and transition to nap. Observers took running notes for each activity and transition.[38] Following the PQA protocol, observers provided evidence for each item rating. The PQA observation period include a teacher interview lasting approximately 20–25 minutes (regarding planning and classroom items). The PQA Form B is collected through an interview with the program director and was completed at the same time as the spring director interview developed for the study.

We also conducted observations in center- and home-based settings using the CLASS Toddler or CLASS Pre-K, depending on the age of children in the classroom or group observed. The CLASS tool focuses on adult-child interactions and instruction practices. The CLASS Toddler has two subscales—Emotional and Behavioral Support and Engaged Support for Learning—and the CLASS Pre-K has three subscales—Emotional Support, Classroom Organization, and Instructional Support. When the CLASS was used to observe classrooms (or groups) with a mix of age levels, observers used both the Toddler and Pre-K CLASS versions, alternating use of each instrument between cycles—the approach recommended by Teachstone (Vitiello, 2014). As directed by the CLASS protocol, observers took notes during as many as six 20-minute observation cycles and immediately rated the dimensions within 10 minutes before beginning the next observation cycle. Use of CLASS Pre-K is not applicable during outdoor free play. Use of CLASS Toddler is applicable during outdoor free play. The rating scale for each CLASS dimension ranges from 1 to 7 (with 7 indicating highest quality), and the overall subscale scores are calculated as the average across all observation cycles.

Finally, CLASS observers completed the CIS on the same day as CLASS observations. The CIS contains 26 items that measure a specific caregiver's interactions with a group of children, unlike the PQA and CLASS tools, which focus on adult-child interactions for the group or classroom setting as a whole. Observers completed the CIS questionnaire for all home-based adults or center-employed adults (e.g., lead teacher, assistant teacher, teacher aides, etc.) assigned to work with the children and present in the group or classroom for at least 45 minutes. For example, we did not rate volunteers, interns, therapists, or special guest speakers with the CIS. Each item on the CIS is scored from 1 to 4, and the 26 items make up four subscales: Sensitivity, Harshness, Detachment, and Permissiveness. Final CIS scores represent the overall

[38] Observers took at least 20 minutes of running notes when applicable during longer activities (such as 45-minute "outdoor play") before pausing to rate items in the PQA protocol.

quality of a caregiver's interactions with the children observed, with higher scores indicating higher-quality interactions.

Other Data Collection Instruments

We include facsimiles of the following additional data collection instruments:

- the Parent Consent Form and Survey, which was completed by a parent to obtain basic child and family background characteristics
- the Information About Children Form covering program enrollment and information about rostered children, which was completed by the provider with the expectation that the provider would be most knowledgeable about the child's program participation and receipt of child care subsidies
- the fall 2014 and spring 2015 director interview questionnaire.

Evaluation of Delaware Stars
Parental Consent Form and Survey

Dear Parent or Guardian:

As described in the enclosed brochure, the Delaware Office of Early Learning has hired the RAND Corporation, a non-profit research organization, to evaluate the state's child care quality rating and improvement system named *Delaware Stars*. The director of your child's program has agreed to participate in the study. We invite you and your child to participate as well. The evaluation includes developmental assessments of ***children ages 2-to-5-years*** in fall 2014 and spring 2015. Results from this study will be used to help improve Delaware's education system.

If you agree to participate:

1. Trained RAND staff may assess your child one-on-one at your child care center. Assessments will be in a "game format." They will measure children's language, literacy, early math skills, and social and emotional skills.
2. Assessments may be administered in two separate sessions as needed for some children. Test time per child is up to 45 minutes.
3. The director of your child's program may share with RAND basic information about your child including: birthdate (or age), gender, date of program enrollment, weekly hours in care, attendance, and receipt of child care subsidy.
4. The child's teacher will fill out a short questionnaire about the child's social and emotional skills.

Children who complete the assessments will be offered a free book for their participation.

Participation: Your decision to take part in the RAND study is voluntary. You and/or your child may withdraw consent at any time without consequence. RAND may not assess any child without parental consent. Children with consent will be randomly selected to participate. RAND will not assess your child or request information about your child if you do not sign and return this consent form. There are no direct benefits or risks associated with this study.

Confidentiality: All of your information, your child's information, and your child's performance on assessments is confidential. It *will not* be shared with anyone outside of the RAND research team, including your child care provider. Individual answers and names will not be included in any study reports.

If you have questions regarding the study, please contact Melissa Bradley, Survey Coordinator, at 1-866-484-7621 (toll free). Information about the study is available at: http://www.rand.org/education/projects/delaware-stars.html and in the enclosed brochure.

Please complete this one-page form if you consent to participate.
Then seal it in the enclosed envelope, and return it to your child's child care or early learning provider.

Consent: By signing this form below, I give permission for my child to be part of the RAND Evaluation of Delaware Stars. I grant RAND permission to assess my child in fall 2014 and again in spring 2015. I understand that RAND may not elect to assess my child, even if I have signed this consent form. I understand that my child's participation is voluntary and he/she may refuse to participate with no consequence. I also give permission for my child's child care program to share basic information about my child with RAND. This information will solely be used for the research study. I acknowledge that I have read, understand, and agree to all contents of this form.

Child's First and Last Name: _____ _____

Child's Date of Birth (MM/DD/YYYY): _ _ / _ _ / _ _ _ _

Parent/Guardian Signature: _____ Date: _____

Parent/Guardian First and Last Name (PRINT): _____ Relationship to Child: _____

Name of Child's Teacher or Room at Child Care: _____

Best Parent Telephone/E-mail to Reach You with Questions About this Form:
Phone: (_____) _____-_____ E-mail address: _____@_____

Please also fill out the 7 questions on the back of this page. ⇨

Demographic Survey - You do not have to answer any question that you do not wish to answer.

What group or groups describe *your* CHILD'S race or ethnic origin?

(CHECK ALL THAT APPLY)

- ❒ Black/African-American
- ❒ White
- ❒ Latino/Hispanic/Latin American/Spanish
- ❒ Indian/South Asian
- ❒ Chinese
- ❒ Korean
- ❒ Vietnamese
- ❒ Arab
- ❒ Other (SPECIFY _____)

How much school have YOU completed? CHECK THE HIGHEST LEVEL YOU HAVE COMPLETED.

(CHECK ONE)

- ❒ None
- ❒ 1-11 years
- ❒ High school graduate/GED
- ❒ Some vocational school
- ❒ Some college
- ❒ Associates' degree (AA)
- ❒ Bachelors' degree (BA, BS)
- ❒ Some graduate or professional school (after completing college)
- ❒ Completed graduate/professional degree

How much school has your child's OTHER PARENT completed? CHECK THE HIGHEST LEVEL COMPLETED.

(CHECK ONE)

- ❒ N/A - I Don't Know
- ❒ None
- ❒ 1-11 years
- ❒ High school graduate/GED
- ❒ Some vocational school
- ❒ Some college
- ❒ Associates' degree (AA)
- ❒ Bachelors' degree (BA, BS)
- ❒ Some graduate or professional school (after completing college)
- ❒ Completed graduate/professional degree

What is your family's income? PLEASE CHECK THE CATEGORY THAT INCLUDES THE TOTAL AMOUNT YOU AND ANY OTHER MEMBERS OF YOUR HOUSEHOLD RECEIVED LAST YEAR IN WAGES, SALARIES, COMMISSIONS, AND TIPS.

(CHECK ONE)

- ❒ $1 - $5,000
- ❒ $5,001 - $10,000
- ❒ $10,001 - $25,000
- ❒ $25,001 - $50,000
- ❒ $50,001 - $100,000
- ❒ $100,001 - $250,000
- ❒ More than $250,000

What is the main language your CHILD speaks at home?

(CHECK ONE)

- ❒ English
- ❒ Spanish
- ❒ Chinese
- ❒ Korean
- ❒ Vietnamese
- ❒ Arabic
- ❒ Other (SPECIFY_____)

How many people live in your household, including yourself?

(CHECK ONE)

- ❒ 2 people
- ❒ 3 people
- ❒ 4 people
- ❒ 5 people
- ❒ 6 people
- ❒ 7 people
- ❒ 8 people
- ❒ Other (SPECIFY) _____

What is your current zip code?

Information about Children in the RAND Evaluation of Delaware Stars

Dear Program Director/Provider:

Please complete the following information about the current number of children enrolled in your center/program and about each child for whom RAND has obtained a signed parental consent form to participate in the evaluation. *Please place the completed form in the envelope provided.* If you have any questions, please contact Melissa Bradley, Survey Director at 703-413-1100, extension 5614.

1. **These questions are about the current number of children enrolled by full day and part day status. By full day, we mean children staying for the morning and the afternoon. By part day, we mean children attending for only a morning session or only an afternoon session but not both. Answer zero for a question if you don't serve the age range or full/part day arrangement.**

Infants [ages 0 to 11 months]	
a. As of today, how many infants [ages 0 to 11 months], if any, are enrolled in your program **part-day for 1-3 days per week**?	**Number:** ☐☐☐
b. As of today, how many infants [ages 0 to 11 months], if any, are enrolled in your program **part-day for 4-5 days per week**?	**Number:** ☐☐☐
c. As of today, how many infants [ages 0 to 11 months], if any, are enrolled in your program **full day for 1-3 days per week**?	**Number:** ☐☐☐
d. As of today, how many infants [ages 0 to 11 months], if any, are enrolled in your program **full day for 4-5 days per week**?	**Number:** ☐☐☐
Toddlers [ages 12 to 36 months]	
e. As of today, how many toddlers [ages 12 to 36 months], if any, are enrolled in your program **part-day for 1-3 days per week**?	**Number:** ☐☐☐
f. As of today, how many toddlers [ages 12 to 36 months], if any, are enrolled in your program **part-day for 4-5 days per week**?	**Number:** ☐☐☐
g. As of today, how many toddlers [ages 12 to 36 months], if any, are enrolled in your program **full day for 1-3 days per week**?	**Number:** ☐☐☐
h. As of today, how many toddlers [ages 12 to 36 months], if any, are enrolled in your program **full day for 4-5 days per week**?	**Number:** ☐☐☐
Preschoolers [ages 3 to 5 years]	
i. As of today, how many preschoolers [ages 3 to 5 years], if any, are enrolled in your program **part-day for 1-3 days per week**?	**Number:** ☐☐☐
j. As of today, how many preschoolers [ages 3 to 5 years], if any, are enrolled in your program **part-day for 4-5 days per week**?	**Number:** ☐☐☐
k. As of today, how many preschoolers [ages 3 to 5 years], if any, are enrolled in your program **full day for 1-3 days per week**?	**Number:** ☐☐☐
l. As of today, how many preschoolers [ages 3 to 5 years], if any, are enrolled in your program **full day for 4-5 days per week**?	**Number:** ☐☐☐
School-age children [kindergarten and older]	
m. As of today, how many school-age children, in any, are enrolled in your program **part-day for 1-3 days per week**?	**Number:** ☐☐☐
n. As of today, how many school-age children, if any, are enrolled in your program **part-day for 4-5 days per week**?	**Number:** ☐☐☐
o. As of today, how many school-age children, if any, are enrolled in your program **full day for 1-3 days per week**?	**Number:** ☐☐☐
p. As of today, how many school-age children, if any, are enrolled in your program **full day for 4-5 days per week**?	**Number:** ☐☐☐

2. Please fill in the table(s) on the following page about each child for whom RAND has obtained a signed parental consent form. Please complete for all listed children on the next 3 pages.

	CHILD 01	CHILD 02	CHILD 03	CHILD 04	CHILD 05
Child Name					
a. What gender is this child?	*Check one:* ☐ Male ☐ Female	*Check one:* ☐ Male ☐ Female	*Check one:* ☐ Male ☐ Female	*Check one:* ☐ Male ☐ Female	*Check one:* ☐ Male ☐ Female
b. What is the date of birth for this child?	____/____/____ MM/DD/YYYY	____/_____/____ MM/DD/YYYY	____/_____/_____ MM/DD/YYYY	____/_____/_____ MM/DD/YYYY	____/_____/_____ MM/DD/YYYY
c. What month and year did this child first enroll in your program?	____/_____ MM/YYYY	____/_____ MM/YYYY	____/_____ MM/YYYY	____/_____ MM/YYYY	____/_____ MM/YYYY
d. Does this child receive Purchase of Care (POC)?	*Check one:* ☐ No ☐ Yes	*Check one:* ☐ No ☐ Yes	*Check one:* ☐ No ☐ Yes	*Check one:* ☐ No ☐ Yes	*Check one:* ☐ No ☐ Yes
e. Does this child participate in Early Childhood Assistance Program (ECAP)?	*Check one:* ☐ No ☐ Yes	*Check one:* ☐ No ☐ Yes	*Check one:* ☐ No ☐ Yes	*Check one:* ☐ No ☐ Yes	*Check one:* ☐ No ☐ Yes
f. How many hours per week does this child attend in your program?	*Range: 1-60 hours* _____ Hrs	*Range: 1-60 hours* _____ Hrs	*Range: 1-60 hours* _____ Hrs	*Range: 1-60 hours* _____ Hrs	*Range: 1-60 hours* _____ Hrs
g. How many days per week is this child enrolled to attend in your program?	*Range: 1-5 days* _____ Days	*Range: 1-5 days* _____ Days	*Range: 1-5 days* _____ Days	*Range: 1-5 days* _____ Days	*Range: 1-5 days* _____ Days
h. Thinking about the last 4 weeks, how many scheduled days did this child *not* attend?	*Range: 0 - 20 days* _____ Days	*Range: 0 - 20 days* _____ Days	*Range: 0 - 20 days* _____ Days	*Range: 0 - 20 days* _____ Days	*Range: 0 - 20 days* _____ Days
i. What's the number or name of this child's classroom?					
j. What days and times does this child attend your program?					
k. Does this child have a diagnosed disability?	*Check one:* ☐ No ☐ Yes	*Check one:* ☐ No ☐ Yes	*Check one:* ☐ No ☐ Yes	*Check one:* ☐ No ☐ Yes	*Check one:* ☐ No ☐ Yes
l. What home language is spoken by this child?					

RAND Delaware Stars Evaluation

Program Director (or Proxy) Interview
RAND Delaware Stars Evaluation
Program Director (or Proxy) Interview
(Fall 2014)
(Interviewer-Administered)

Verbal consent - Delaware Stars Evaluation; Provider Interviews

I work for the RAND Corporation, which is a non-profit research organization. We've been hired by the Delaware Office of Early Learning to evaluate Delaware Stars, which is your state's quality rating and improvement system for child care and early learning programs. The goal of the study is to improve early child care and education in Delaware overall and to make improvements to the Delaware Stars program where needed.

Today we would like to learn more about your program and get your input on Delaware Stars. We are mainly interviewing directors or program providers. But if a director/provider names an alternative administrator for the interview, we are interviewing him or her instead.

The information you provide will be kept strictly confidential and will not affect your current or future Delaware Stars rating or your access to other Delaware Stars supports such as technical assistance or financial incentive. We will not share your responses with the Delaware Office of Early Learning, or anyone else outside of the project, except as required by law. And, we will not identify any individuals by name in our study reports.

Your participation in this interview is voluntary. You do not have to discuss anything that you do not feel comfortable discussing. You may choose not to participate, decline to answer any question, or stop the interview at any time without any penalty. This interview may take up to [*Fall 2014: 45-minutes; Spring 2015: 30-minutes*). Do you agree to participate in the interview?

[IF NO:] Thank you anyway.

[IF YES:] If you have any questions about the study, I'm handing out to you a paper with information about the study, and the contact information of people that you can call or email with questions about the study.

Do you have any questions before we begin?

Part 1. Validate Program Information

RROLEF **TYPE OF RESPONDENT COMPLETING THE FALL INTERVIEW**
<1> CHILD CARE PROVIDER, PROGRAM DIRECTOR, OWNER
<2> CURRICULUM COORDINATOR
<3> OTHER, SPECIFY:
<d> DON'T KNOW
<r> REFUSED

INTRO: **Okay, I want to first make sure that some of the data we have out your program is up to date and accurate.**

LICSTART **In what month did this program first get a license to start operation of this facility?**
MONTH:
YEAR:
<d> DON'T KNOW
<r> REFUSED

PROVIDER **Is your program a:**
<1> Licensed family child care facility (LCC)
<2> Licensed day care center (Early Care and Education) (LDC)
<3> Licensed large family child care center (LLF), or
<4> Something else? SPECIFY:
<d> DON'T KNOW
<r> REFUSED

PROPR **How is the operator of this center (the person or organization who runs the center) organized? A child care center may be classified as a nonprofit by the state or federal government, or the center may be considered for-profit if it is operated by an individual, group, or company which has the potential to earn profits from providing the services.** ***(IF NEEDED: This refers to the main or primary operator.)***
Is it:
<1> For profit, [GO TO PROPR2]
<2> Not-For-Profit, [GO TO PROPR3]
<3> A public agency, or [GO TO PROPR4]
<4> something else? SPECIFY: [GO TO PARAGRAPH BEFORE DSTART]
<d> DON'T KNOW [GO TO PARAGRAPH BEFORE DSTART]
<r> REFUSED [GO TO PARAGRAPH BEFORE DSTART]

PROPR2 **What specific type of For-Profit is this center? Is it:**
<1> An independent owner/operator,
<2> A local or regional chain of two or more centers,
<3> A national chain which franchises to a local operator,
<4> A national chain which operates the center directly, or
<5> An on-site center operated by a business for its employees?
<6> Other
<d> DON'T KNOW
<r> REFUSED
[ALL GO TO PARAGRAPH BEFORE DSTART]

PROPR3 **What specific type of Not-For-Profit is this center? Is it:**

<1> A parent cooperative
<2> A private grade or high school or college
<3> A church
<4> Affiliated with, but not operated by, a church
<5> An independent nonprofit center, but not one of the above
<6> A nongovernmental community agency
<7> Other
<d> DON'T KNOW
<r> REFUSED

[ALL GO TO PARAGRAPH BEFORE DSTART]

PROPR4 **What specific type of Public Agency is this center? Is it:**

<1> Public elementary, middle, or high school
<2> Federal agency, Head Start
<3> DECC Early Childhood Assistance Program (state agency)
<4> Public college or university
<5> State or local government agency, but not one of the above
<6> Federal agency, but not Head Start
<7> Other
<d> DON'T KNOW
<r> REFUSED

INTRO **The next six questions are about the director/provider's education and years of experience. (IF RROLEF NOT EQUAL TO 1: If you are not the director/provider, I will either need to obtain this information from you about the director/provider, or else obtain this directly from the director/provider after the interview.)**

DSTART **What date did you begin work as the director/provider of this program?**

MONTH: [MM]
YEAR: [YYYY]
<d> DON'T KNOW
<r> REFUSED

DYRSKIDS **How many years of experience do you have directly working with children either as a director/provider or a teacher in a child care program or preschool?**

TOTAL NUMBER OF YEARS: _____
<d> DON'T KNOW
<r> REFUSED

DYRSDIR **How many total years of experience do you have a director/provider of a child care program or preschool?**

TOTAL NUMBER OF YEARS: _____
<d> DON'T KNOW
<r> REFUSED

DED **Which, if any, degrees do you have? (Do you have a:)**

ENTER 1 FOR ALL THAT APPLY; WRITE IN FIELD OF STUDY AS APPLICABLE FOR EACH DEGREE SELECTED

<1> High School diploma or GED,

<1> Associate degree,

AA_FIELD. Field of Study (write in) SPECIFY:

<1> Bachelor's degree,

BA_FIELD. Field of Study (write in) SPECIFY:

<1> Master's degree,

MA_FIELD. Field of Study (write in) SPECIFY: DED@maos

<1> Doctorate degree (Ph.D., Ed.D.), or

DR FIELD. Field of Study (write in) SPECIFY:

<1> Something else?

SPECIFY:

<d> DON'T KNOW

<r> REFUSED

DCRED **Which, if any, credentials or qualifications in early childhood education do you have? (Do you have a:)**

(SELECT ALL THAT APPLY)

<1> Child Development Associate (CDA) Credential

<1> Delaware DOE Early Childhood Assistant Teacher Qualification

<1> Delaware DOE Early Childhood Teacher Qualification

<1> Delaware DOE Early Childhood Curriculum Coordinator (with or without degree)

<1> Delaware DOE Early Childhood Administrator Qualification

<1> Delaware DOE School-Age Administrator Qualification

<1> Delaware DOE Infant-Toddler Care Credential

<1> Delaware DOE School-Age Care Credential

<1> Delaware DOE Program Administration Credential

<1> Some other early childhood credential

<1> Or no early childhood credential

<d> DON'T KNOW

<r> REFUSED

DTRAIN **Which, if any, best describes any education or training the director/provider has had to prepare for the job as director/provider? (Choose one that best describes the director/provider.)**

<1> No particular courses, on the job training

<2> The required coursework for compliance with state regulations

<3> Required coursework plus some workshops

<4> Coursework in administration, management, leadership, adult education or a related field

<5> Coursework toward a degree, or completion of a degree in administration, adult education or related field

<d> DON'T KNOW

<r> REFUSED

NAEYC **Is your program accredited by the National Association for the Education of Young Children (NAEYC)?**

<1> YES

<5> NO

<d> DON'T KNOW

<r> REFUSED

HSTART **Does your program receive funding through the federal Head Start or Early Head Start programs?**
<1> YES
<5> NO [GO TO ECAP]
<d> DON'T KNOW [GO TO ECAP]
<r> REFUSED [GO TO ECAP]

HSTARTENR **How many Head Start children are currently enrolled in your program? By enrollment, I'm referring to the unique number of children of any age that you serve in Head Start regardless of whether they are full time or part time.**
NUMBER: _____
<d> DON'T KNOW
<r> REFUSED

ECAP **Does your program receive funding through the Delaware Early Childhood Assistance Program (ECAP)?**
<1> YES
<5> NO [GO TO POC]
<d> DON'T KNOW [GO TO POC]
<r> REFUSED [GO TO POC]

ECAPENR **How many children are enrolled in ECAP? By enrollment, I'm referring to the unique number of children of any age that you serve regardless of whether they are full time or part time.**
NUMBER: _____
<d> DON'T KNOW
<r> REFUSED

POC **Does your program accept Purchase of Care payments? If so, what kind?**
IF NEEDED: Is that Purchase of Care or Purchase of Care **Plus**?
<1> YES, PURCHASE OF CARE ACCEPTED
<2> YES, PURCHASE OF CARE **PLUS** ACCEPTED
<3> NO, PURCHASE OF CARE NOT ACCEPTED [GO TO PART]
<4> OTHER. SPECIFY: [GO TO PART]
<d> DON'T KNOW [GO TO PART]
<r> REFUSED [GO TO PART]

POCENR **How many children receiving purchase of care are enrolled as of today? By enrollment, I'm referring to the unique number of children of any age that you serve regardless of whether they are full time or part time.**
NUMBER: ____
<d> DON'T KNOW
<r> REFUSED

PART **Is your program enrolled in Delaware Stars for Early Success, the state's quality rating and improvement system for child care and early learning programs?**
<1> YES
<5> NO [GO TO ROOMS]
<d> DON'T KNOW [GO TO ROOMS]
<r> REFUSED [GO TO ROOMS]

ENR **What month and year did your program first enroll in Delaware Stars?**

MONTH: _____
YEAR: _____
<d> DON'T KNOW
<r> REFUSED

ETYLEV **What was your star level when you first enrolled?**

NUMBER: _____ [1-5] [IF ETYLEV=1 GOTO LVL]
<d> DON'T KNOW [GO TO LVL]
<r> REFUSED [GO TO LVL]

ALT **Are you certified by an alternative path because you are a public school, Head Start, NAEYC certified, ECAP, or some other reason?**

<1> YES
<5> NO
<d> DON'T KNOW
<r> REFUSED

LVL **What is your star level currently?**

NUMBER: _____ [1-5]
<d> DON'T KNOW
<r> REFUSED

VALID **What month and year was your program last validated by the Delaware Stars assessors?**

MONTH: [MM]
YEAR: [YYYY]
<d> DON'T KNOW
<r> REFUSED

ROOMS **How many classes or rooms of children do you have in this program?**

NUMBER: _____
<d> DON'T KNOW
<r> REFUSED

ANSCHED **What months of year, if any, is your program not open for school/child care? Please indicate week ranges if your program is closed for a portion of the month. For example, closed for 3rd & 4th week of May.**

SELECT 1 FOR ALL THAT APPLY AND THEN SPECIFY WEEK RANGES
<1> January, SPECIFY WEEK RANGES:
<1> February, SPECIFY WEEK RANGES:
<1> March, SPECIFY WEEK RANGES:
<1> April, SPECIFY WEEK RANGES:
<1> May, SPECIFY WEEK RANGES:
<1> June, SPECIFY WEEK RANGES:
<1> July, SPECIFY WEEK RANGES:
<1> August, SPECIFY WEEK RANGES:
<1> September, SPECIFY WEEK RANGES:
<1> October, SPECIFY WEEK RANGES:
<1> November, SPECIFY WEEK RANGES:
<1> December, SPECIFY WEEK RANGES:

SPEDENR **How many identified special needs children are enrolled at present? I am looking for a headcount. By special needs we mean children with either a physical disability (including hearing or sight problems) mental disabilities, or emotional disabilities. (Identified means identified by parents and program staff, and possibly, but not necessarily, by an outside agency).**

NUMBER: _____ 00-999

RACE **Approximately what percentage of children enrolled in your program come from the following ethnic or racial groups?** ***(Write in percentages—e.g., 99 if you mean 99%.)***

White, non Hispanic: _____
African-American/Black, non Hispanic: _____
Hispanic/Latino: _____
Asian/Pacific Islander: _____
Other: _____

ONEN **Approximately what percentage of children enrolled in your program currently primarily speak a language other than English?** ***(Write in percentages—e.g., 99 if you mean 99%.)***

PERCENT: _____

Part 2. Communication with Public about Delaware Stars

INTRO **Now I want to ask you about how parents perceive Delaware Stars. This is the final section of the interview today.**

INFORM **Do you inform parents interested in enrolling their child in your program of your Delaware Stars rating?**

<1> YES
<5> NO [GO TO NOADVERT]
<d> DON'T KNOW [GO TO NOADVERT]
<r> REFUSED [GO TO NOADVERT]

ADVERT **What methods do you use to inform prospective parents about your rating? Choose as many as apply.**

<1> Include printed outreach materials provided by Office of Early Learning in application Packet [GO TO ASK]
<1> Explain verbally to applicants [GO TO ASK]
<1> Hang up posters, signs, or banners in the Program [GO TO ASK]
<1> Refer applicants to Stars website [GO TO ASK]
<1> Something else? SPECIFY: [GO TO ASK]
<1> Something else? SPECIFY: [GO TO ASK]

NOADVERT **Why do you not advertise to applicants?**

_____

_____

ASK **Do applicants ask you if your program is Delaware Stars rated? Would you say:**

<1> Yes,
<2> No, applicants rarely, if ever, ask, or [GO TO PERCEIVE]
<3> something else? SPECIFY: ASK@os [GO TO PERCEIVE]
<d> DON'T KNOW [GO TO PERCEIVE]
<r> REFUSED [GO TO PERCEIVE]

INTRO **I have two follow up questions about that.**

ASKPCT **What percent of applicants ask in a typical month?**
PERCENT: _____

ASKNUM **What's the total number of applicants who ask about Delaware Stars in a typical month?**
NUMBER: _____

PERCEIVE **Do you think applicants perceive Delaware Stars as a good gauge of quality? Why or why not?**

_____

_____

APPLICANT **Do you think you get more applicants at your program because of your Delaware Stars rating?**
<1> YES
<5> NO
<d> DON'T KNOW
<r> REFUSED

END. **Thank you for your time!**

RAND Delaware Stars Evaluation

Program Director (or Proxy) Interview
(Spring 2015)
(Interviewer-Administered)

RROLESPR Type of respondent completing the fall interview
- 1 Child care provider, program director, owner
- 2 Curriculum coordinator
- 3 Other: _____ 30 CHARACTERS

Verbal consent - Delaware Stars Evaluation; Provider Interviews

I work for the RAND Corporation, which is a non-profit research organization. We've been hired by the Delaware Office of Early Learning to evaluate Delaware Stars, which is your state's quality rating and improvement system for child care and early learning programs. The goal of the study is to improve early child care and education in Delaware overall and to make improvements to the Delaware Stars program where needed.

Today we would like to learn more about your program and get your input on Delaware Stars. We are mainly interviewing directors or program providers. But if a director/provider names an alternative administrator for the interview, we are interviewing him or her instead.

The information you provide will be kept strictly confidential and will not affect your current or future Delaware Stars rating or your access to other Delaware Stars supports such as technical assistance or financial incentive. We will not share your responses with the Delaware Office of Early Learning, or anyone else outside of the project, except as required by law. And, we will not identify any individuals by name in our study reports.

Your participation in this interview is voluntary. You do not have to discuss anything that you do not feel comfortable discussing. You may choose not to participate, decline to answer any question, or stop the interview at any time without any penalty. This interview may take up to [*Fall 2014: 45-minutes; Spring 2015: 30-minutes*). Do you agree to participate in the interview?

[IF NO:] Thank you anyway.

[IF YES:] If you have any questions about the study, I'm handing out to you a paper with information about the study, and the contact information of people that you can call or email with questions about the study.

Do you have any questions before we begin?

Part 1. Validate Program Information

[Interviewer: Ask questions LICSTART - DTRAIN only if A or B is true:

- ***(A) Is the FALL 2014 INTERVIEW.***
- **(B) It is the SPRING 2015 interview and there is some missing information from FALL 2014. If this is the case, ask only the questions from the insert indicating missing information.**

Part 1. Section A

Okay, I want to first make sure that some of the data we have out your program is up to date and accurate. [Spring interview only: Some of these questions are repeated from the fall, but we just want to make sure our information is up to date.]

LICSTART In what month did this program first get a license to start operation of this facility?
month: _____ 01-12
year: _____ 1900-2014

PROVIDER Is your program a:
1 LCC. licensed family child care facility,
2 LDC. licensed day care center (Early Care and Education)
3 LLF. licensed large family child care center
4 Other. _____ SPECIFY

PROPR How is the operator of this center (the person or organization who runs the center) organized? A child care center may be classified as a nonprofit by the state or federal government, or the center may be considered for-profit if it is operated by an individual, group, or company which has the potential to earn profits from providing the services. *(Note: This refers to the main or primary operator.)*

Is it:

For profit 01

Not for profit 02

A public agency 03

Other. _____ SPECIFY 04

For Profit Is it:

An independent owner/operator 01

A local or regional chain of two or more centers 02

A national chain which franchises to a local operator 03

A national chain which operates the center directly 04

An on-site center operated by a business for its employees 05

Not-For-Profit Is it:

A parent cooperative 06

A private grade or high school or college 07

A church 08

Affiliated with, but not operated by, a church 09

An independent nonprofit center, but not one of the above 10

A nongovernmental community agency 11

Other 12

IF OTHER SPECIFY:

As a Public Agency Is it:

Public elementary, middle, or high school 13

Federal agency, Head Start 14

DECC Early Childhood Assistance Program (state agency) 15

Public college or university 16

State or local government agency, but not one of the above 17

Federal agency, but not Head Start 18

Other 19

IF OTHER SPECIFY:

The next six questions are about the director/provider's education and years of experience. If you are not the director/provider, I will either need to obtain this information from you about the director/provider, or else obtain this directly from the director/provider after the interview.

DSTART What date did you begin work as the director/provider of this program?
MONTH: _____ [MM] 01-12
YEAR: _____ [YYYY] 1900-2014

DYRSKIDS How many years experience do you have directly working with children either as a director/provider or a teacher in a child care program or preschool?

Total number of years: _____ yrs. 00-99

DYRSDIR How many total years experience do you have a director/provider of a child care program or preschool?
Total number of years: _____ yrs. 00-99

DED Which, if any, degrees do you have?
(Select all that apply; write in field of study as applicable for each degree selected)
High School diploma/GED 1
Associate degree 2
AA_FIELD. Field of Study (write in)_____ SPECIFY
Bachelor's degree 3
BA_FIELD. Field of Study (write in)_____ SPECIFY
Master's degree 4
MA_FIELD. Field of Study (write in)_____ SPECIFY
Doctorate degree (Ph.D., Ed.D.) 5
DR_FIELD. Field of Study (write in) _____SPECIFY
Other SPECIFY_____ 6

DCRED Which, if any, credentials/qualifications in early childhood education do you have? (Select all that apply)

Child Development Associate (CDA) Credential 01

Delaware DOE Early Childhood Assistant Teacher Qualification 02

Delaware DOE Early Childhood Teacher Qualification 03

Delaware DOE Early Childhood Curriculum Coordinator

(with or without degree) 04

Delaware DOE Early Childhood Administrator Qualification 05

Delaware DOE School-Age Administrator Qualification 06

Delaware DOE Infant-Toddler Care Credential 07

Delaware DOE School-Age Care Credential 08

Delaware DOE Program Administration Credential 09

Other early childhood credential 10

No early childhood credential 11

DTRAIN Which, if any, best describes any education or training the director/provider has had to prepare for the job as director/provider? (Choose one that best describes the director/provider.)

No particular courses, on the job training 1

The required coursework for compliance with state regulations 2

Required coursework plus some workshops 3

Coursework in administration, management, leadership, adult education or a related field 4

Coursework toward a degree, or completion of a degree in administration,

adult education or related field 5

Part I. Section B

Interviewer: Ask these questions in both FALL 2014 and in SPRING 2015 interviews.

NAEYC Is your program accredited by the National Association for the Education of Young Children (NAEYC)?

1 Yes

5 No

HSTART Does your program receive funding through the federal Head Start or Early Head Start programs?

1 Yes

5 No = GOTO ECAP

<d,r> = GOTO ECAP

HSTARTENR How many Head Start children are currently enrolled in your program? By enrollment, I'm referring to the unique number of children of any age that you serve in Head Start regardless of whether they are full time or part time.

_____ 00-999

ECAP Does your program receive funding through the Delaware Early Childhood Assistance Program (ECAP)?

1 Yes

5 No = GOTO POC

<d,r> = GOTO POC

ECAPENR How many children are enrolled in ECAP? By enrollment, I'm referring to the unique number of children of any age that you serve regardless of whether they are full time or part time.

_____ 00-999

POC Does your program accept Purchase of Care payments? If so, what kind?

1 Yes, Purchase of Care accepted

2 Yes, Purchase of Care **Plus** accepted

3 No, Purchase of Care not accepted = GOTO PART

4 Other. _____ SPECIFY= GOTO PART

<d,r> = GOTO PART

POCENR How many children receiving purchase of care are enrolled as of today? By enrollment, I'm referring to the unique number of children of any age that you serve regardless of whether they are full time or part time.

____ 00-999

PART Is your program enrolled in Delaware Stars for Early Success, the state's quality rating and improvement system for child care and early learning programs?

1 Yes

5 No <d,r> = GOTO SPEDENR

<d,r> = GOTO SPEDENR

[SKIP ENR if SPRING 2015:]

ENR What month and year did your program first enroll in Delaware Stars?

MONTH: _____ [MM] 01-12

YEAR: _____ [YYYY] 1900-2014

[SKIP ETYLEV IF SPRING 2015:]

[If PART = yes]

ETYLEV What was your star level when you first enrolled?

_____ [1-5]

IF ETYLEV=1 GOTO LVL

<d,r> = GOTO LVL

ALT Are you certified by an alternative path because you are a public school, Head Start, NAEYC certified, ECAP, or some other reason?

1 Yes

5 No

LVL What is your star level currently?

_____ [1-5]

VALID What month and year was your program last validated by the Delaware Stars assessors?

MONTH: _____ [MM] 01-12

YEAR: _____ [YYYY] 1900-2014

STARPLUS Is your program in the Stars Plus Cohort that gets extra TA support?

1 Yes

5 No

[if STARPLUS = Yes:]

TAPLUS Do you think the extra support provided for being in the Stars Plus Cohort has helped improve the quality of your program? Why or why not?

OPEN ENDED – NO FIELD LIMIT

_____

_____

[Skip ROOMS if SPRING 2015:]

ROOMS How many classes or rooms of children do you have in this program? _____ 00-99

[Skip ANSCHED if SPRING 2015:]

ANSCHED What months of year, if any, is your program **not** open for school/child care? Select all the apply, if please indicate week ranges if your program is closed for a portion of the month. For example, closed for 3rd & 4th week of May. SPECIFY

a. January _____ SPECIFY WEEK RANGES
b. February _____ SPECIFY WEEK RANGES
c. March _____ SPECIFY WEEK RANGES
d. April _____ SPECIFY WEEK RANGES
e. May _____ SPECIFY WEEK RANGES
f. June _____ SPECIFY WEEK RANGES
g. July _____ SPECIFY WEEK RANGES
h. August _____ SPECIFY WEEK RANGES
i. September _____ SPECIFY WEEK RANGES
j. October _____ SPECIFY WEEK RANGES
k. November _____ SPECIFY WEEK RANGES
l. December _____ SPECIFY WEEK RANGES

>INFENR_PDPW<. As of today, how many infants [ages 0 to 11 months], if any, are enrolled in your program part-day for 1-3 days per week?

ENTER NUMBER: @

[@] <0-999, d, r>

INFENR_PDAW.. As of today, how many infants [ages 0 to 11 months], if any, are enrolled in your program part-day for 4-5 days per week? _____ 00-999

I INFENR_FDPW. As of today, how many infants [ages 0 to 11 months], if any, are enrolled in your program full day for 1-3 days per week? _____ 00-999

INFENR_FDAW. As of today, how many infants [ages 0 to 11 months], if any, are enrolled in your program full day for 4-5 days per week? _____ 00-999

TODENR_PDPW. As of today, how many toddlers [ages 12 to 36 months], if any, are enrolled in your program part-day for 1-3 days per week? _____ 00-999

TODENR_PDAW. As of today, how many toddlers [ages 12 to 36 months], if any, are enrolled in your program part-day for 4-5 days per week? _____ 00-999

TODENR_FDPW. As of today, how many toddlers [ages 12 to 36 months], if any, are enrolled in your program full day for 1-3 days per week? _____ 00-999

TODENR_FDAW. As of today, how many toddlers [ages 12 to 36 months], if any, are enrolled in your program full day for 4-5 days per week? _____ 00-999

PREKENR_PDPW. As of today, how many preschoolers [ages 3 to 5 years], if any, are enrolled in your program part-day for 1-3 days per week?? _____ 00-999

PREKENR_PDAW As of today, how many preschoolers [ages 3 to 5 years], if any, are enrolled in your program part-day for 4-5 days per week? _____ 00-999

PREKENR_FDPW. As of today, how many preschoolers [ages 3 to 5 years], if any, are enrolled in your program full day for 1-3 days per week? _____ 00-999

PREKENR_FDAW. As of today, how many preschoolers [ages 3 to 5 years], if any, are enrolled in your program full day for 4-5 days per week? _____ 00-999

SAGEENR_PDPW. As of today, how many school-age children, in any, are enrolled in your program part-day for 1-3 days per week? Include kindergarten or older. _____ 00-999

SAGEENR_PDAW. As of today, how many school-age children, if any, are enrolled in your program part-day for 4-5 days per week? _____ 00-999

SAGEENR_FDPW. As of today, how many school-age children, if any, are enrolled in your program full day for 1-3 days per week? _____ 00-999

SAGEENR_FDAW. As of today, how many school-age children, if any, are enrolled in your program full day for 4-5 days per week? _____ 00-999

SPEDENR How many identified special needs children are enrolled at present? I am looking for a headcount. By special needs we mean children with either a physical disability (including hearing or sight problems) mental disabilities, or emotional disabilities. (Identified means identified by parents and program staff, and possibly, but not necessarily, by an outside agency). _____ 00-999

Approximately what percentage of children enrolled in your program come from the following ethnic or racial groups? *(Write in percentages—e.g., 99 if you mean 99%.)*

White, non Hispanic:	_____	[PCT_WHT] 00-99
African-American/Black, non Hispanic	_____	[PCT_AA] 00-99
Hispanic/Latino:	_____	[PCT_HISP] 00-99
Asian/Pacific Islander:	_____	[PCT_API] 00-99
Other:	_____	[PCT_OTHER] 00-99

ONEN Approximately what percentage of children enrolled in your program currently primarily speak a language other than English? *(Write in percentages—e.g., 99 if you mean 99%.)*

_____ 00-99

Interviewer: Proceed to Part 2 if this is the fall Interview & Delaware Stars provider = yes. Otherwise, this is the end of the FALL Interview Providers (FALL = yes & DE Star Participate = No). *This is the end of our interview for this fall. Thank you for your time today!*

Part 2. Communication with Public about Delaware Stars

(A) Interviewer directions: Part 2 applies only if the program currently participates in Delaware Stars.

Now I want to ask you about how parents perceive Delaware Stars.

INFORM Do you inform parents interested in enrolling their child in your program of your Delaware Stars rating?

1 Yes
5 No =GOTO NOADVERT

<d,r>=GOTO NOADVERT

ADVERT What methods do you use to inform prospective parents about your rating? Choose as many as apply.

1 Include printed outreach materials provided by Office of Early Learning in application packet=GOTO ASK
2 Explain verbally to applicants=GOTO ASK
3 Hang up posters, signs, or banners in the program=GOTO ASK
4 Refer applicants to Stars website=GOTO ASK
5 Other: _____ 30 CHARACTERS=GOTO ASK
6 Other: _____ 30 CHARACTERS=GOTO ASK

NOADVERT Why do you not advertise to applicants?
OPEN ENDED – NO FIELD LIMIT

_____

_____

ASK Do applicants ask you if your program is Delaware Stars rated?

1 Yes
2 No, applicants rarely, if ever, ask =GOTO PERCEIVE
3 Other. _____ 30 CHARACTERS =GOTO PERCEIVE

<d,r>=GOTO PERCEIVE

[SKIP ASKPCT AND ASKNUM IF SPRING 2015:]

I have two follow up questions about that.

ASKPCT What percent of applicants ask in a typical month?

_____ 00-99

ASKNUM What's the total number of applicants ask about Delaware Stars in a typical month?

_____ 00-999

PERCEIVE Do you think applicants perceive Delaware Stars as a good gauge of quality? Why or why not?
OPEN ENDED – NO FIELD LIMIT

_____

_____

APPLICANT Do you think you get more applicants at your program because of your Delaware Stars rating?

1 Yes
5 No

Part 3. Provider perceptions of Delaware Stars

Interviewer directions: **Part 3 applies only if this is the spring 2015 interview**.

For DESTARS PARTICIPANTS ONLY (PART = yes), complete following Qs in Part 3, Section A. For NON-DESTARS PARTICIPANTS, SKIP TO Part 3, Section B.

Interviewer directions*: When asking Part 3 questions, please pull out copies of each of the 3 Delaware Stars 100-point standards documents (1 document each for LFCC, FCC, ECE providers).*

Part 3, Section A

We're in the final section of the interview. In this section my questions are about your perceptions of what does and doesn't work in Delaware Stars. First, I have four questions about the program as a whole. Then I'll follow with questions about specific elements of Stars.

REASON Why did you decide to participate in Delaware Stars? Please select as many answers as apply.
I'll read all the answer options to you, and then please tell me all that apply.
FIRST CODE ALL THAT APPLY AND THEN RANK THEM

Interviewer: **If <u>2 or more</u> benefits selected say:** Rank your answers in the space provided; 1 = Most beneficial

1 I wanted to improve the quality of my program.
2 I wanted more professional recognition
3 I wanted to make my child care or preschool program more attractive to parents.
4 I wanted new ideas for my child care or preschool program.
5 I wanted access to the grants and other financial incentives available via Delaware Stars.
6 I wanted the technical assistance that Delaware Stars offers.
7 I wanted to attract and retain qualified staff.
8 I wanted to increase my business.
9 Other: _____ (fill in) SPECIFY

[Source: Elicker, Indiana Paths of Quality evaluation]

VALUE What aspects, if any, of Delaware Stars have been most beneficial to you?
I'll read all the answer options to you, and then please tell me all that apply.
FIRST CODE ALL THAT APPLY AND THEN RANK THEM

Interviewer: **If <u>2 or more</u> benefits selected say:** Rank your answers in the space provided; 1 = Most beneficial

1 The technical assistance. ____
2 The grants and financial incentives.____
3 The recognition I get from parents, other providers, or that public that I am providing high quality care.____
4 Participation provides me with a marketing tool for my child care or preschool program.____
5 The Environmental Rating Scores for my classrooms.____
6 The validation process (including rating on individual standards and the ERS).____
7 Other benefit not listed here: _____SPECIFY
8 No benefits.

[Source: Elicker, Indiana Paths of Quality evaluation]

MOVEUP Are you actively planning to move up to the next Star level?

1 Yes; I have applied or I am planning to apply to move up another level.
2 No; I have no plans to apply to move up another Star level.
3 Not applicable. Already rated as Star 5; no higher level.
4 Other: _____ SPECIFY

[Source: Elicker, Indiana Paths of Quality evaluation]

NO_MOVEUP. [If MOVEUP = NO.] Why do you not have plans to apply to move up another Star level?

_____

_____

BARR In your opinion, what has been the biggest obstacle to moving up to the next Star level? I'll read the potential answers to you. You may select all that apply.

FIRST CODE ALL THAT APPLY AND THEN RANK THEM

Interviewer: **If 2 or more obstacles selected say:** Rank your answers in the space provided; 1 = Biggest obstacle

1 Finding the time to complete tasks required for the next level
2 Completion of required staff education and training.
3 Insufficient funding to meet standards.
4 Getting the paperwork and documentation in order.
5 Preparing for and meeting national accreditation standards.
6 Having to wait 6 months to get the next ERS assessment.
7 Preparing for and meeting the required ERS score
8 Need more feedback from technical assistance provider
9 Challenges in developing a curriculum.
10 Other: __BARR.10@os____ SPECIFY
11 Not applicable. Rated as star 5 at **entry**; no higher level.
12 No obstacles.

[Source: Elicker, Indiana Paths of Quality evaluation]

Now I have questions about specific aspects of Delaware Stars.

[if *ALT=YES GOTO STDSHARD]*

VALIDATE Thinking back to your most recent validation by a Delaware Stars assessor, how useful was the observation and validation of paperwork? Did the assessor identify useful ways to improve program quality? Why or why not?

_____

_____

STDSHARD Are there particular standards in the Delaware Stars system that you view as especially **difficult** to attain? If so, which ones and why? (NOTE: *this question is not referring to ERS standards. We ask about that separately.*)

_____

_____

STDSEASY Are there particular standards in the Delaware Stars system that you view as **easy** to attain? If so, which ones and why? (NOTE: *this question is not referring to ERS standards. We ask about that separately.*)

_____

_____

STDSBAG Are there particular standards in the Delaware Stars system that you view as **not contributing to quality** and should thus be excluded from the ratings? If so, which ones and why? (NOTE: *this question is not referring to ERS standards. We ask about that separately.*)

_____

_____

TAVALUE Does your Delaware Stars technical assistance (TA) provider offer guidance and feedback that you view as improving your program?

_____

_____

TAIMP How do you think Delaware Stars technical assistance (TA) could be improved, if at all?

_____

_____

TASPEC Have you ever received any of the following types of specialized technical assistance? Mark all that apply.

1 early childhood mental health consultation,
2 child assessment,
3 Ages & Stages developmental screening,
4 ERS specialist,
5 infant/toddler care,
6 health and nutrition
NONE OF THE ABOVE,
<d,r> = GOTO A4E

TASPECMPR Do you think that specialized technical assistance has helped to improve your program? Which ones, and why or why not?

_____

_____

A4E [Have you / Has the director of this program] participated in the Early Learning Leadership Initiative Aim4Excellence course?

1 Yes
5 No = GOTO ERSIMP
3 Not applicable
<d,r> = GOTO ERSIMP

A4EIMP Do you think that participation in Aim4Excellence has helped to improve your program? Why or why not?

_____

_____

ERSIMP How do you think Delaware Stars Environmental Rating Score process could be improved, if at all? Explain.

_____

_____

Now I want to ask you about each type of financial incentives that Delaware Stars offers to see if you have applied for it, ever received it, or view it as useful.

INFRA *Infrastructure Fund: These funds became available as of late summer 2013 to programs at a Star 2 or higher for capital improvements & technology needs.*
Have you ever obtained an infrastructure grant award for either technology or for infrastructure?

1 Yes
5 No = GOTO CORE
3 Not applicable
<d,r> = GOTO CORE

[If INFRA = yes]
INFRAUSE Do you view this grant as being potentially useful to improving the quality of your program? Why or why not?

_____

_____

CORE *Compensation, Retention, and Education (CORE) Awards are for professionals who meet specified educational requirements.*
Do any of your staff currently have or have they ever received any type of CORE award?

1 Yes
5 No = GOTO QI
3 Not applicable
<d,r> = GOTO QI

COREUSE Do you view this grant as being potentially useful to improve the quality of your program? Why or why not?

_____

_____

QI *Quality Improvement grants for materials and professional development for programs at Star 2-Star 4. These range from up to $750 for family child care providers up to $5,000 for the largest centers.*
Have you ever obtained one or more materials & professional development grants (now called quality improvement grants)?

1 Yes
5 No= GOTO POCGR
3 Not applicable = GO TO POCGR
<d,r> = GOTO POCGR

QIPDUSE Do you view this grant as being potentially useful to improve the quality of your program? Why or why not?

_____

_____

POCGR *Tiered reimbursement for Purchase of Care. Providers with a 3, 4, of 5 star rating get 80%, 93%, or 102% of the market rate respectively for children who are eligible for Purchase of Care. The funds are distributed on a quarterly basis.*

Has your program ever received tiered reimbursement for Purchase of Care?

1 Yes
5 No = GOTO SCHOLAR
3 Not applicable = GO TO SCHOLAR
<d,r> = GOTO SCHOLAR

POCGRUSE Do you view this tiered reimbursement as an effective tool to improve the quality of your program? Why or why not?

_____

_____

SCHOLAR *T.E.A.C.H. Early Childhood® scholarships. Staff in Delaware Stars programs have priority status for T.E.A.C.H. scholarships that can be used to pay for early childhood teachers and administrators pursuing a degree or credential related to early childhood education at a participating Delaware higher education institution.*

Have one or more of your staff received or currently have a TEACH scholarship?

1 Yes
5 No
3 Don't know

FUNDING Thinking about all the possible financial incentives available in Delaware Stars—POC tiered reimbursement, quality improvement grants, infrastructure fund, CORE awards, and T.E.A.C.H. Early childhood scholarships—which one do you think is the most important to you? (choose one only)

1 POC tiered reimbursement,
2 quality improvement grants,
3 infrastructure fund,
4 CORE awards, or
5 T.E.A.C.H. Early childhood scholarships
6 OTHER (specify): _____ FUNDING@os

STAFF *Okay, my final question is about the impact that Stars has had, if any, on your staff.*

Can you talk about the impact Delaware Stars has had, if any, on your staff—particularly teachers? Have you seen a change in your teachers? If yes, how so? If not, why do you think there is no change?

_____

_____

END. Thank you for your time!

Part 3, Section B

Interviewer directions: These only apply if this is the spring 2015 interview and the program does NOT currently participate in Delaware Stars.

Okay, I have two last questions for you.

NONPART What are the main reasons you do not participate in Delaware Stars currently? I'll read answers to you, and please select all that apply.
FIRST CODE ALL THAT APPLY AND THEN RANK THEM

Interviewer: **If <u>2 or more</u> benefits selected say:** Rank your answers in the space provided; 1 = Main reason

1 Too much time / too burdensome to enroll
2 Too expensive to meet standards
3 Not an effective marketing tool to attract applicants
4 Not a good measure of program quality
5 I plan to, but I haven't done it yet.
6 Don't know
7 Other: _____ SPECIFY
8 Other: _____ SPECIFY

MAYPART Are there changes the state could make to the Delaware Stars system that would cause you to participate in the program? If so, what would those changes include?

_____

_____

END. Thank you for your time!

Appendix D. Additional Characteristics of the Provider Sample

To supplement information presented in Chapter 2, this appendix presents additional characteristics of the programs in the provider sample (Table D.1) and the characteristics of the respondents to the provider survey (Table D.2). These survey-based indicators are not otherwise used in the analyses presented in Chapters 3 and 4. All results are unweighted.

Table D.1. Characteristics of Programs in Delaware Stars: Director interview Responses

Characteristic	Unweighted Percentage
Year first enrolled in Delaware Stars (percentage distribution)	
Enrolled 2006 to 2010	17.1
Enrolled 2011 to 2012	32.6
Enrolled 2013 to 2014	25.4
Unknown	24.9
Year rating was last validated (percentage distribution)	
2010 or earlier	0.6
2011 or 2012	2.8
2013	12.2
2014	49.2
2015	22.1
Unknown	13.3
Number of programs	181

SOURCE: Authors' tabulations of provider sample director survey.

Table D.2. Characteristics of Director Interview Respondents

Characteristic	Unweighted Percentage
Role (percentage distribution)	
Director or owner	92.8
Other role	6.6
Unknown	0.6
Respondent years of tenure with program (percentage distribution)	
Fewer than 5 years	5.5
5 years to less than 10 years	13.8
10 years to less than 20 years	37.6
20 years or more	42.5
Unknown	0.6
Respondent years of experience as a director (percentage distribution)	
Fewer than 5 years	23.8
5 years to less than 10 years	33.7
10 years to less than 20 years	29.3
20 years or more	12.7
Unknown	0.6
Respondent highest education (percentage distribution)	
High school diploma or GED	20.4
Associate degree	24.9
Bachelor's degree	35.4
Master's degree or higher	18.2
Unknown	1.1
Credentials (percentage distribution)	
Has one or more ECE credential	93.4
Has no ECE credential or don't know	6.6
Specific credentials (percentage; more than one may apply)	
Has CDA credential	13.8
Has Delaware DOE Early Childhood Assistant Teacher Qualification	30.9
Has Delaware DOE Early Childhood Teacher Qualification	49.7
Has Delaware DOE Early Childhood Curriculum Coordinator (with or without degree)	37.0
Has Delaware DOE Early Childhood Administrator Qualification	63.0
Has Delaware DOE School-Age Administrator Qualification	33.7
Has Delaware DOE Infant-Toddler Care Credential	35.9
Has Delaware DOE School-Age Care Credential	31.5
Has Delaware DOE Program Administrator Credential	37.6
Has other early childhood credential	23.8
Training received to prepare for director position (percentage distribution)	
No particular courses, on-the-job training	6.1
Required coursework for compliance with state regulations	16.0
Required coursework plus some workshops	26.0
Coursework in administration, management, leadership, adult education, or a related field	19.9
Coursework toward a degree in or completion of a degree in administration, adult education, or a related field	30.4
Unknown	1.7
Number of programs	181

SOURCE: Authors' tabulations of provider sample director interview.
NOTE: CDA = Child Development Associate. DOE = Department of Education.

Appendix E. Statistical Methods

This appendix provides additional information about the methods used in the analyses of the provider sample and child sample data presented in Chapters 3 and 4. Additional sensitivity analyses beyond the preferred models are described in Appendix G.

Chapter 3 Preferred Model Specification

To estimate the relationship between alternative measures of ECE program quality and Delaware Stars ratings (as well as components of Delaware Stars ratings), we specified a series of regressions for the analyses summarized in Chapter 3. Specifically, our main regression models took the following form:

$$Y_{jk} = \alpha_j X_j + \beta_j P_j + \delta_k N_k + \varepsilon_{jk},$$

where Y_{jk} is the program-level winter 2015 score for a given alternative measure of quality in program j, located in neighborhood k; X_j is the main predictor of interest, usually a program-level variable, such as the Delaware Stars rating or ratings component (e.g., ERS score, score in each of the four quality domains); P_j is the vector of program-level covariates; N_k is the vector of provider neighborhood-level covariates; and $\varepsilon_{jk}$are provider-level errors. All models were linear.

To assess the statistical significance of the predictor of interest (e.g., Delaware Stars rating status), we first performed a joint test of significance of the categories of the predictor (e.g., Starting with Stars or Star 2, Star 3, Star 4, or Star 5). If that test indicated that there was a statistically significant relationship between the predictor and the outcome, we also then performed pairwise tests between categories of the predictor to determine which pairs differed. We first applied the standard cutoff of $\alpha = 0.05$ for statistical significance. However, because we performed many tests, it is possible that some significant findings are due to chance. To control for false positives—i.e., finding statistically significant differences by chance—we implemented the Benjamini-Hochberg procedure to adjust for multiple hypothesis testing (Benjamini and Hochberg, 1995). The Benjamini-Hochberg procedure essentially applies a stricter standard for statistical significance. In Chapter 3 and in the appendixes, we present both the unadjusted and adjusted statistical significance of results. However, we prioritize the corrected results when discussing the findings and their implications for policy.

To aid in the interpretability of results, we presented in Chapter 3 bar charts of "adjusted means" of outcomes at each level of the primary variable of interest—e.g., average classroom observation ratings in programs rated Starting with Stars or Star 2, Star 3, Star 4, and Star 5. These means are adjusted for all program and program neighborhood characteristics used as covariates in the model by standardizing the population of children receiving each level of the main predictor of interest to all children in the model (Graubard and Korn, 1999).

Table E.1 provides descriptive statistics for the alternative measures of quality and the covariates included in the provider-level models reported in Chapter 3.

Table E.1. Descriptive Statistics for Provider Sample Analyses of Quality and Delaware Stars Ratings

Characteristic	Weighted Mean	Standard Deviation
Alternative measures of quality		
PQA	3.33	1.17
CLASS Pre-K Emotional Support	5.32	1.85
CLASS Pre-K Classroom Organization	4.57	2.33
CLASS Pre-K Instructional Support	2.14	1.36
CLASS Toddler Emotional and Behavioral Support	5.18	2.15
CLASS Toddler Engaged Support for Learning	2.65	1.88
CIS	3.26	0.88
Provider type		
Small FCC	34.7	–
Large FCC	4.0	–
Center	61.3	–
School-based		
Delaware Stars rating as of August 2015 (percentage distribution)		
Not in Stars	34.0	–
Starting with Stars	2.3	–
Star 2	18.2	–
Star 3	10.4	–
Star 4	19.2	–
Star 5	15.8	–
Time in months between spring 2015 observation and effective date of August 2015 Delaware Stars rating		
For Starting with Stars/Star 2	2.38	13.44
For Star 3 to Star 5	3.30	13.01
Part B Section 619 indicator	0.8	–
Title 1 indicator	1.5	–
Head Start indicator	3.3	–
ECAP indicator	1.3	–
NAEYC accreditation indicator	3.9	–
If verified, last done under prior standards		–
If verified, last done under prior standards	71.5	–
No	28.5	–
Provider zip code characteristics		
Race-ethnicity of population (average percentage distribution)		
Non-Hispanic white	57.3	43.9
Non-Hispanic black	28.0	42.0
Hispanic	9.3	14.6
Non-Hispanic other	5.4	5.6
Education of adults age 15 and above (average percentage distribution)		
Less than high school	13.4	12.6
High school diploma or GED	32.5	12.1
Some college	27.2	6.4
Bachelor's degree or higher	26.8	23.5
Average percentage of family households with children under 18 and headed by single parent	18.6	20.8
Average percentage of children under age six living in families with income below the federal poverty line	13.4	15.5
Average median household income (thousands of dollars)	60.2	43.5
N	181	

SOURCES: Authors' tabulations of licensed database and provider sample disposition file, matched to U.S. Census zip code characteristics. State administrative data are the source for Delaware Stars participation and license type.
NOTES: Percentage distributions are calculated excluding missing cases. Percentage distributions may not sum to 100 because of rounding. – = not applicable.

Outcomes

We used three alternative measures of ECE program quality: PQA, CLASS, and CIS, described further in Appendix C. Because Delaware Stars ratings are at the program level, in programs where we observed multiple classrooms, we first created a program-level average score for each of the alternative measures of classroom quality that we collected (e.g., PQA, CLASS scales). For the CIS, we first created a classroom average across all staff with a CIS score and then created a program-level average across the classroom averages. Alternative specifications, described in Appendix G, used the minimum score across the observed classrooms within each program (or staff within each classroom).

Main Predictors of Interest

We used Delaware Stars status (in Stars or not in Stars) and rating tier as of August 2015. We combined Starting with Stars and Star 2 as one level for two reasons. The first is that we assessed very few programs (and few children) from Starting with Stars programs, and the second reason is that ECE programs at Starting with Stars or Star 2 have not undergone formal assessments (of either the ERS classroom observations or standards-based validation). Thus, both types of programs can be viewed as "in process." Our final parameterizations of Delaware Star status were not in Delaware Stars, Starting with Stars or Star 2, Star 3, Star 4, and Star 5 (see Table E.1).

We also estimated models based on components of the Delaware Stars ratings—namely, the points-based standards, the essential standards, and the ERS. These models could be estimated only for programs at Star 3 and higher. For the points-based standards, we estimated separate models for each of the four domains: Qualifications and Professional Development, Family and Community Partnerships, Management and Administration, and Learning Environment and Curriculum. Scoring for each of these domains varies by program type (FCC versus center). Moreover, the method of scoring changed somewhat in October 2014 (see the discussion in Chapter 1), and some but not all programs had been assessed under the new set of standards as of August 2015. In order to analyze points-based scores produced under all scoring methods in a single model, we first converted scores into the same scale (mean = 0, standard deviation = 1) within each of four strata (old standards FCC, old standards center, new standards FCC, and new standards center) by subtracting stratum-specific means and dividing by stratum-specific standard deviations. We then divided the standardized scores into four categories based on quartiles of the combined distribution.

OEL has designated six standards as "essential standards": annual child developmental screening; twice annual formative child assessment; written comprehensive curriculum; daily activities, lesson plans, and individualized goal plans; administrator possesses certain credentials; and curriculum coordinator possesses the Curriculum and Assessment Credential. Not all six standards apply to FCCs. We parameterized each of these six essential standards as a yes/no indicator to indicate whether the program had any points versus no points on that essential standard at the time of its last rating, and we modeled each standard separately. We also used the

sum of the points that the program obtained on the six essential standards (thereby treating the essential standards as a bundle) as a predictor in a separate model. We again adjusted for differences in scoring by program type (centers and FCCs) and for old versus new standards using quartiles of the combined standardized scores as cut points to create four categories, as was done for the four quality domains.

The ERS was also standardized within each of the same four strata, and the combined standardized scores were used to create four classifications based on quartiles.

Provider Controls

To allow for possible differences in the relationship between Delaware Stars ratings and observed quality depending on provider type and programs that can be rated through an alternative pathway, we controlled for these variables. We coded provider type as small FCC, large FCC, small center (fewer than 50 enrollees), and large center (50 or more enrollees), based on licensing data. Because we assessed very few children in large FCCs, this provider type was grouped with small centers for most models, except for models of the components that make up the ratings (i.e., quality domain scores, essential standards met, and ERS score) because requirements for centers and FCCs are different. Other provider covariates were selected to account for the providers eligible for an alternative pathway (see Table E.1).

Provider Neighborhood–Level Controls

We collected programs' current zip code from OCCL and Delaware Stars administrative data and confirmed with directors. We then cleaned and linked provider zip codes to five-year (2009–2013) ACS data. Table E.1 reports the zip code–level variables that were included as controls. For a few programs, we could not match the program to a zip code with ACS measures.

Chapter 4 Preferred Model Specification

We followed a similar approach to estimating the relationship between Delaware Stars ratings (or ratings components) and child developmental outcomes, with results presented in Chapter 4. Our main regression models took the following form:

$$Y_{ijk} = \alpha_j X_j + \beta_j P_j + \gamma_i C_i + \delta_k N_k + \varepsilon_{ijk},$$

where Y_{ijk} is the spring 2015 score for child i, attending program j, and residing in neighborhood k; X_j is the main predictor of interest, usually a program-level variable like a Delaware Stars rating, but sometimes a classroom-level variable; P_j is the vector of program-level covariates; C_i is the vector of child-level covariates, including the child's score on the same assessment in fall 2014; N_k is the vector of child neighborhood-level covariates; and $\varepsilon_{ijk}$are child-level errors. Because we assessed up to 15 children per program, we adjusted the standard errors of our estimates by clustering them at the provider level using Taylor series linearization. All models were linear models.

As with the models described above for Chapter 3, we employed a similar strategy for hypothesis testing, including adjusting for multiple hypothesis testing. We also present regression-adjusted average child assessment scores by levels of the predictor of interest. We further divided the adjusted means by the sample standard deviation of the outcome variable. This means that the difference between adjusted means of, for example, children's scores in Starting with Stars or Star 2 versus Star 3 programs are reported in effect sizes, which can be interpreted as small (0.2), medium (0.5), and large (0.8) (Cohen, 1992, 1998). As we describe in Chapter 4, we did not find any statistically significant differences that fall in the range of effect sizes deemed medium or large.

Table E.2 presents descriptive statistics for the child developmental outcomes and the provider- and child-level covariates included in the models. The models were estimated using the 1,123 children assessed in fall 2014 and spring 2015 (column B). We provide a discussion of attrition from the child sample at the end of this section.

Outcomes

As described in Appendix C, we administered five developmental assessments: PPVT, WJ–Letter Word Identification, WJ–Applied Problems, HTKS, and DECA (which produces two subscales). HTKS produces only raw scores that are the sum of correct answers. But all the others yield several types of scores, and we used those that technical documentation indicated were appropriate for measuring growth in skills from one time point to another. Specifically, we used growth scale values (GSV) for the PPVT, w-scores for the WJ subscales, raw total scores for HTKS, and t-scores for the DECA. Because our primary interest was measuring growth within a cohort of children over time, we did not use age- or grade-normed scores for assessments where they were available (e.g., for the WJ subscales and PPVT).

Main Predictors of Interest

As with the models for Chapter 3, we were primarily interested in the relationship between Delaware Stars rating (as of August 2015)—or the components that make up the rating—and child developmental outcomes. For those measures, we followed the same approach described above.

Provider- and Child-Level Controls

We included the provider-level covariates listed in Table E.1 (see Table E.2). The same ACS-based zip code characteristics were used, except they were based on the zip code reported on the parent survey that accompanied the consent form (see Table E.2). In addition to the characteristics listed in Table E.2, the models included a control for the child's fall assessment score for the assessment being modeled, as well as the number of months between the fall and spring assessments.

Table E.2. Provider and Child Characteristics for Children Assessed in Fall 2014, Assessed in Fall 2014 and Spring 2015, and Who Attrited

Controls	Children Assessed in Fall 2014 (A)		Children Assessed in Fall 2014 and Spring 2015 (B)		Children Who Attrited (C)		
Child assessment in fall 2014 (unstandardized)							
PPVT	112.2	(21.7)	112.9	(21.7)	107.7	(21.3)	**
WJ–Applied Problems	392.6	(25.3)	393.6	(25.1)	387.2	(25.7)	**
WJ–Letter Word Identification	320.7	(30.5)	321.9	(30.5)	314.0	(29.7)	***
HTKS	8.1	(13.4)	8.5	(13.6)	5.7	(11.9)	*
DECA–Total Protective Factors	43.1	(6.7)	43.2	(6.8)	42.7	(6.7)	
DECA–Absence of Behavioral Problems	32.6	(4.2)	36.6	(4.1)	32.8	(4.5)	
Provider-level characteristics							
Provider type							***
Large center	976	(74.2%)	852	(75.9%)	124	(64.6%)	
Small center	321	(24.4%)	253	(22.5%)	68	(35.4%)	
Large FCC	18	(1.4%)	18	(1.6%)	0	(0.0%)	
Time in months between spring 2015 observation and effective date of August 2015 Delaware Stars rating							
For Starting with Stars/Star 2			1.8	(6.4)			–
For Star 3 to Star 5			5.7	(8.6)			–
Part B Section 619 indicator	36	(2.7%)	35	(3.1%)	1	(0.5%)	*
Title 1 indicator	51	(3.9%)	47	(4.2%)	4	(2.1%)	
Head Start indicator	167	(12.7%)	139	(12.4%)	28	(14.6%)	
ECAP indicator	34	(2.6%)	25	(2.2%)	9	(4.7%)	*
NAEYC accreditation indicator	83	(6.3%)	81	(7.2%)	2	(1.1%)	**
If verified, last done under prior standards							
No	288	(40.1%)	253	(40.4%)	35	(38.0%)	
Yes	431	(59.9%)	374	(59.7%)	57	(62.0%)	
Child- and family-level characteristics							
Child's gender							*
Female	680	(51.7%)	567	(50.5%)	113	(58.9%)	
Male	635	(48.3%)	556	(49.5%)	79	(41.2%)	
Child's race-ethnicity							***
Non-Hispanic white	564	(44.1%)	508	(46.2%)	56	(31.1%)	
Non-Hispanic black	356	(27.8%)	271	(24.7%)	85	(47.2%)	
Hispanic	211	(16.5%)	186	(16.9%)	25	(13.9%)	
Non-Hispanic other	148	(11.6%)	134	(12.2%)	14	(7.8%)	
% Missing	*36*	*(2.7%)*	*24*	*(2.1%)*	*12*	*(6.3%)*	
Family income							***
$10,000 or less	178	(16.0%)	136	(14.2%)	42	(27.3%)	
$10,001–$25,000	238	(21.4%)	192	(20.1%)	46	(29.9%)	
$25,001–$50,000	198	(17.8%)	172	(18.0%)	26	(16.9%)	
$50,001–$100,000	229	(20.6%)	206	(21.5%)	23	(14.9%)	
$100,001 or more	268	(24.1%)	251	(26.2%)	17	(11.0%)	
Missing	*204*	*(15.5%)*	*166*	*(14.8%)*	*38*	*(19.8%)*	
Higher of parents' education attainment							***
High school graduate/GED or less	176	(14.2%)	144	(13.5%)	32	(18.6%)	
Some vocational school, some college, or associate's degree	512	(41.3%)	410	(38.4%)	102	(59.3%)	
Bachelor's degree or higher	551	(44.5%)	513	(48.1%)	38	(22.1%)	
Missing	*76*	*(5.8%)*	*56*	*(5.0%)*	*20*	*(10.4%)*	

Table E.2. Provider and Child Characteristics for Children Assessed in Fall 2014, Assessed in Fall 2014 and Spring 2015, and Who Attrited, Continued

Controls	Children Assessed in Fall 2014 (A)		Children Assessed in Fall 2014 and Spring 2015 (B)		Children Who Attrited (C)		
School year child first enrolled in ECE program							***
August 2013 or earlier	511	(40.2%)	463	(42.4%)	48	(26.5%)	
September 2013 to August 2014	393	(30.9%)	336	(30.8%)	57	(31.5%)	
September 2014 or later	368	(28.9%)	292	(26.8%)	76	(42.0%)	
Missing	*43*	*(3.3%)*	*32*	*(2.8%)*	*11*	*(5.7%)*	
POC recipient							***
No	848	(66.6%)	759	(69.6%)	89	(48.9%)	
Yes	425	(33.4%)	332	(30.4%)	93	(51.1%)	
Missing	*42*	*(3.2%)*	*32*	*(2.8%)*	*10*	*(5.2%)*	
ECAP recipient							***
No	1,170	(92.3%)	1,007	(92.9%)	163	(88.6%)	
Yes	98	(7.7%)	77	(7.1%)	21	(11.4%)	
Missing	*47*	*(3.6%)*	*39*	*(3.5%)*	*8*	*(4.2%)*	
Hours per week child attends the program							
Less than 12.5	113	(8.8%)	101	(9.2%)	12	(6.5%)	
12.5 to less than 20	60	(4.7%)	55	(5%)	5	(2.7%)	
20 to less than 40	332	(25.9%)	277	(25.3%)	55	(29.6%)	
40 or more	778	(60.6%)	664	(60.5%)	114	(61.3%)	
Missing	*32*	*(2.4%)*	*6*	*(2.3%)*	*26*	*(3.1%)*	
Chronic absence (missed 5 or more days in last 4 weeks)							
No	1,169	(91.3%)	1,004	(91.6%)	165	(89.2%)	
Yes	112	(8.7%)	92	(8.4%)	20	(10.8%)	
Missing	*34*	*(2.6%)*	*27*	*(2.4%)*	*7*	*(3.6%)*	
Child has a diagnosed disability							
No	1,215	(95.7%)	1,043	(95.7%)	172	(95.6%)	
Yes	55	(4.3%)	47	(4.3%)	8	(4.4%)	
Missing	*45*	*(3.4%)*	*33*	*(2.9%)*	*12*	*(6.3%)*	
Child's home language and language of assessment							
English (home), English (assessed)	1,123	(85.4%)	952	(84.8%)	171	(89.1%)	
Spanish or Spanish/English (home), English (assessed)	58	(4.4%)	52	(4.6%)	6	(3.1%)	
Other language or other language/English (home), English (assessed)	48	(3.7%)	42	(3.7%)	6	(3.1%)	
Characteristics of child's home zip code							
Race-ethnicity of population (average percentage distribution)							
Non-Hispanic white	62.8	(19.6)	63.5	(19.2)	58.0	(21.7)	***
Non-Hispanic black	22.5	(18.1)	21.7	(17.5)	27.0	(20.8)	***
Hispanic	9.5	(7.4)	9.4	(7.3)	10.3	(7.8)	
Non-Hispanic other	5.2	(2.9)	5.3	(3.0)	4.7	(2.6)	***
Education of adults age 15 and above (average percentage distribution)							
Less than high school	13.7	(6.5)	13.5	(6.5)	15.4	(6.7)	***
High school diploma or GED	32.7	(7.4)	32.4	(7.5)	34.7	(5.9)	***
Some college	26.9	(3.6)	26.8	(3.7)	27.1	(3.3)	
Bachelor's degree or higher	26.6	(13.4)	27.3	(13.7)	22.7	(10.4)	***

Table E.2. Provider and Child Characteristics for Children Assessed in Fall 2014, Assessed in Fall 2014 and Spring 2015, and Who Attrited, Continued

Controls	Children Assessed in Fall 2014 (A)	Children Assessed in Fall 2014 and Spring 2015 (B)	Children Who Attrited (C)	
Average percentage of family households with children under age 18 and headed by single parent	16.6 (8.6)	16.1 (8.3)	19.3 (10.2)	***
Average percentage of children under age six living in families with income below the federal poverty line	22.3 (14.2)	21.6 14.0)	26.5 (14.9)	***
Average median household income (thousands of dollars)	61.2 (21.1)	62.3 (21.5)	54.7 (16.4)	***
Missing zip code	*49 (3.7%)*	*34 (3.0%)*	*15 (7.8%)*	
Number of children	1,315	1,123	192	

SOURCES: Authors' analysis of provider sample, Delaware Stars administrative data, and child sample.
NOTES: Counts and percentages (in parentheses) are reported for provider and child characteristics. Percentage distributions exclude missing cases and may not sum to 100 because of rounding. The percentage of missing cases is shown for each measure for reference. Means and standard deviations (in parentheses) are reported for child assessment scores and zip-code–level characteristics. Two additional child-level predictors used in models (months between assessments and age at spring 2015 assessment) varied between model outcomes and are not shown. Statistical significances of test for equality of distribution in column (B) versus column (C) are as follows: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. – = not applicable.

Missing Data

We had some missing data in the child-level control variables, either in instances where parents or directors left survey items blank, administrative measures were missing, or we had missing data on the measures linked by zip code. We wished to retain in our models each child who was assessed in both fall 2014 and spring 2015, both because this increases our power to detect differences in children's performance across, say, higher- versus lower-level Star programs and because there may be systematic differences between children who have complete data and those who do not. Therefore, we created a missing data indicator for each covariate and included those in the main specification. Rates of missing data were generally low (see Table D.3). Among the child and family controls, the highest rate of missing data was for household income: nearly 15 percent. This derived from the parent/guardian questionnaire, and some parents skipped this question when consenting for their child to participate in the study. All other covariates were missing for between 0 and 5 percent of cases. Three percent of parents/guardians did not report their home zip code on the parent survey; neighborhood controls are missing for these cases.

Analysis of Child Sample Attrition Between Fall 2014 and Spring 2015

Of the 1,315 children who were assessed in fall 2014, 1,123 (85.4 percent) were assessed again in spring 2015. The remaining 192 children who were assessed in fall 2014 attrited from the study before spring 2015. Directors of participating programs reported that virtually all of the 192 children were no longer enrolled in the programs as of spring 2015, although a small number of children did not provide verbal assent to the spring assessment at the time of administration. Distributions of provider, child, family, and neighborhood controls are shown in Table E.2 for

three sets of children: all children assessed in the fall, children assessed at both time points, and children who attrited from the study. Only data from the 1,123 children with assessments measured at both time points were used in modeling.

We tested whether covariate distributions for the 192 children who attrited between fall 2014 and spring 2015 were different than for the 1,123 children who had data in both waves, using chi-squared tests for categorical variables and t-tests for continuous variables. We found several differences at the 0.05 significance level. Specifically, attriters were more likely than nonattriters to attend small centers where fewer than 50 children enrolled than to attend large centers with more than 50 enrollees or large FCCs. Attriters were more likely than nonattriters to attend ECAP programs and less likely to attend Title 1 or NAEYC-certified programs. Attriters had lower mean fall 2014 scores than nonattriters for four of the assessments: PPVT, WJ–Applied Problems, WJ–Letter Word Identification, and HTKS. Attriters and nonattriters also differed on average for several child- and family-level characteristics: Attriters were more likely to be female, to be non-Hispanic black, to live in households with lower income, to have parents with lower educational attainment, to have enrolled in the day care program more recently, to have POC subsidies, and to have ECAP. Attriters and nonattriters were also different, on average, for several neighborhood-level variables that reflect differences found at the child and family levels: Attriters were from neighborhoods with lower average educational attainment, lower median income, higher poverty rates among children under age five, higher rates of households with children headed by a single parent, and higher percentages of residents who were non-Hispanic black.

Appendix F. Additional Documentation of Analyses

This appendix provides tables that document results presented in Chapters 3 and 4.

Supplemental Tables for Chapter 3

Table F.1. Regression-Adjusted Average Score of Alternative Quality Measures by Quartiles of Quality Domain Scores and ERS Score in Delaware Stars Rating

	Quartile of Quality Domain Score or ERS Score			
Alternative Program Quality Measure	**Lowest**	**Second-Lowest**	**Second-Highest**	**Highest**
Family and Community Partnerships				
PQA	3.48	3.42	3.35[d]	3.59[c]
CLASS Pre-K Emotional Support	5.70	5.78	5.27[d]	5.78[c]
CLASS Pre-K Classroom Organization	4.98	4.84	4.40	4.96
CLASS Pre-K Instructional Support	2.28	2.26	2.02	2.37
CLASS Toddler Emotional and Behavioral Support	5.21	5.45	5.49	5.72
CLASS Toddler Engaged Support for Learning	2.88	2.86	2.98	3.12
CIS	3.32	3.45[C]	3.17[B,d]	3.38[c]
Qualifications and Professional Development				
PQA	3.31	3.46	3.46	3.54
CLASS Pre-K Emotional Support	5.42	5.57	5.86	5.64
CLASS Pre-K Classroom Organization	4.44[c]	4.64	5.11[a]	4.79
CLASS Pre-K Instructional Support	1.73[C,D]	2.27	2.38[A]	2.45[A]
CLASS Toddler Emotional and Behavioral Support	4.92[c,d]	5.72	5.83[a]	5.59[a]
CLASS Toddler Engaged Support for Learning	2.36[c,D]	3.04	3.35[a]	3.14[A]
CIS	3.27	3.31	3.42	3.33
Management and Administration				
PQA	3.43	3.31	3.49	3.49
CLASS Pre-K Emotional Support	6.07[B,c,d]	5.42[A]	5.38[a]	5.65[a]
CLASS Pre-K Classroom Organization	5.18[B]	4.36[A]	4.76	4.71
CLASS Pre-K Instructional Support	2.29[b]	1.79[a,D]	2.01[d]	2.55[B,c]
CLASS Toddler Emotional and Behavioral Support	5.49	5.67	5.21	5.52
CLASS Toddler Engaged Support for Learning	3.04	2.91	2.71	3.07
CIS	3.46	3.31	3.28	3.33
Learning Environment and Curriculum				
PQA	3.28	3.22[C,D]	3.54[B]	3.54[B]
CLASS Pre-K Emotional Support	5.61	5.56	5.98	5.58
CLASS Pre-K Classroom Organization	4.66	4.68	5.28	4.71
CLASS Pre-K Instructional Support	1.82[C,D]	1.85[C,D]	2.60[A,B]	2.35[A,B]
CLASS Toddler Emotional and Behavioral Support	5.33	5.28	5.73	5.49
CLASS Toddler Engaged Support for Learning	2.81	2.60	3.27	3.00
CIS	3.34	3.25[c]	3.49[b]	3.33

Table F.1. Regression-Adjusted Average Score of Alternative Quality Measures by Quartiles of Quality Domain Scores and ERS Score in Delaware Stars Rating, Continued

Alternative Program Quality Measure	Quartile of Quality Domain Score or ERS Score			
	Lowest	Second-Lowest	Second-Highest	Highest
ERS Total Score				
PQA	3.17[B,C]	3.58[A]	3.53[A]	3.42
CLASS Pre-K Emotional Support	5.43	5.83	5.93[d]	5.47[c]
CLASS Pre-K Classroom Organization	4.50	4.95	4.97	4.70
CLASS Pre-K Instructional Support	1.94	2.22	2.36	2.29
CLASS Toddler Emotional and Behavioral Support	4.43[B,C,D]	5.95[A]	5.61[A]	5.33[A]
CLASS Toddler Engaged Support for Learning	2.12[B,C,D]	3.21[A]	2.89[A]	3.07[A]
CIS	3.19[b,c]	3.46[a]	3.45[a]	3.28

SOURCE: Authors' analysis of provider sample classroom observations and Delaware Stars administrative data.
NOTES: Sample restricted to programs at Star 3 to Star 5 excluding those rated through an alternative pathway. Average scores by quartile of the quality domain scores are regression-adjusted from models described in Appendix E. Superscript letters denote that the difference in the regression-adjusted average score is significantly different at $p < 0.05$ prior to adjusting for multiple hypothesis testing from [a] lowest quartile, [b] second-lowest quartile, [c] second-highest quartile, and [d] highest quartile. Capitalized superscript letters denote differences that remain statistically significant after adjusting for multiple hypothesis testing.

Table F.2. Regression-Adjusted Difference in Average Classroom Observation Score for Programs Meeting an Essential Standard Versus Programs Not Meeting the Standard in Delaware Stars Rating

	Essential Standard					
Alternative Program Quality Measure	**Annual Child Developmental Screening**	**Twice-Annual Formative Child Assessment**	**Written Comprehensive Curriculum**	**Daily Activities, Lesson Plans, Individualized Goal Plans**	**Administrator Credentialed**	**Curriculum Coordinator Credential**
PQA	0.24*	–	–	–	–	–
CLASS Pre-K Emotional Support	–	–	–	–	–	–
CLASS Pre-K Classroom Organization	–	–	–	–	–	–
CLASS Pre-K Instructional Support	–	–	–	–	–	–
CLASS Toddler Emotional and Behavioral Support	0.61*	–	–	–	0.75*	–
CLASS Toddler Engaged Support for Learning	0.58*	–	–	–	0.61*	–
CIS	–	–	–	–0.16*	–	–

SOURCE: Authors' analysis of provider sample classroom observations and Delaware Stars administrative data.
NOTES: Sample restricted to programs at Star 3 to Star 5 excluding those rated through an alternative pathway. Difference in average classroom observation score by meeting or not meeting each essential standard is regression-adjusted from models described in Appendix E. A single asterisk (*) denotes that the difference in the regression-adjusted average score for meeting the standard versus not meeting the standard is significantly different at $p < 0.05$ prior to adjusting for multiple hypothesis testing. A double asterisk (**) indicates the difference remains significant after adjusting for multiple hypothesis testing. – = not statistically significant prior to adjusting for multiple hypothesis testing.

Table F.3. Regression-Adjusted Average Score of Alternative Quality Measures by Quartiles of Sum of Points for Essential Standards in Delaware Stars Rating

	Quartile of Sum of Points for Essential Standards			
Alternative Program Quality Measure	**Lowest**	**Second-Lowest**	**Second-Highest**	**Highest**
PQA	3.34	3.45	3.46	3.47
CLASS Pre-K Emotional Support	5.86	5.77	5.61	5.58
CLASS Pre-K Classroom Organization	4.81	5.00	4.96	4.66
CLASS Pre-K Instructional Support	1.96	2.13	2.18	2.34
CLASS Toddler Emotional and Behavioral Support	4.94[c]	5.65	5.68[a]	5.51
CLASS Toddler Engaged Support for Learning	2.60	3.08	2.93	3.07
CIS	3.45	3.37	3.35	3.30

SOURCE: Authors' analysis of provider sample classroom observations and Delaware Stars administrative data.
NOTES: Sample restricted to programs at Star 3 to Star 5 excluding those rated through an alternative pathway. Average scores by quartile of sum of points for essential standards are regression-adjusted from models described in Appendix E. Superscript letters denote that the difference in the regression-adjusted average score is significantly different at $p < 0.05$ prior to adjusting for multiple hypothesis testing from [a] lowest quartile, [b] second-lowest quartile, [c] second-highest quartile, and [d] highest quartile. Capitalized superscript letters denote differences that remain statistically significant after adjusting for multiple hypothesis testing.

Table F.4. Unadjusted Average Score of Alternative Quality Measures by Program Characteristic

Program Characteristic	Number of Programs	Alternative Program Quality Measure						
		PQA	CLASS Pre-K Emotional Support	CLASS Pre-K Classroom Organization	CLASS Pre-K Instructional Support	CLASS Toddler Emotional and Behavioral Support	CLASS Toddler Engaged Support	CIS
Program license type								
Center	132	3.18	5.30	4.48	2.25	5.07	2.76	3.28
Small FCC	34	3.81	5.74	4.94	2.00	5.64	2.67	3.33
Delaware Stars status								
Not in Delaware Stars	13	3.22	5.17	4.58	2.07	4.99	2.58	3.23
In Delaware Stars	156	3.32	5.40	4.57	2.23	5.23	2.73	3.29
Stars Plus cohort member	78	3.26	5.23	4.32	2.13	5.19	2.60	3.22
Rated through alternative pathway								
Head Start stand-alone	14	3.47	5.58	4.45	2.46	–	–	3.40
Accepts POC subsidies	141	3.29	5.32	4.48	2.19	5.18	2.69	3.25
More than 50% enrollees with POC	64	3.19	5.08	4.16	1.89	4.97	2.53	3.14
Enrolls any ECAP children	21	3.41	5.54	4.58	2.31	5.67	2.98	3.35
More than 50% ECAP enrollees	11	3.41	5.54	4.43	2.08	5.67	2.98	3.30

SOURCE: Authors' analysis of provider sample classroom observations and Delaware Stars administrative data.
NOTES: We excluded categories of programs where there were fewer than 10 cases. These categories included large FCCs, ECAP stand-alone programs, NAEYC-accredited programs, and public school Section 619 programs. – = not applicable.

Table F.5. Regression-Adjusted Average Score of Alternative Quality Measures by Program License Type

Alternative Program Quality Measure	Program License Type	
	Centers	Small FCC
PQA	3.15[B]	3.83[A]
CLASS Pre-K Emotional Support	5.34[b]	5.83[a]
CLASS Pre-K Classroom Organization	4.55	5.03
CLASS Pre-K Instructional Support	2.28	2.04
CLASS Toddler Emotional and Behavioral Support	5.00[B]	5.64[A]
CLASS Toddler Engaged Support for Learning	2.75	2.76
CIS	3.29	3.39
Number of programs		
Where PQA observations completed	126	34
Where CLASS Pre-K observations completed	132	32
Where CLASS Toddler observations completed	88	28
Where CIS observations completed	131	31

SOURCE: Authors' analysis of provider sample classroom observations and Delaware Stars administrative data.
NOTES: Average scores by program license type are regression-adjusted from models described in Appendix E. Superscript letters denote that the difference in the regression-adjusted average score is significantly different at $p < 0.05$ prior to adjusting for multiple hypothesis testing from [a] centers and [b] small FCCs. Capitalized superscript letters denote differences that remain statistically significant after adjusting for multiple hypothesis testing.

Table F.6. Regression-Adjusted Difference in Average Classroom Observation Score for Programs With a Characteristic Versus Programs Without the Characteristic

Alternative Program Quality Measure	Program Characteristic			
	Star Plus	Head Start Stand-Alone	A Majority of Children Served Receive POC Subsidies	A Majority of Children Served Receive ECAP
PQA	–0.16*	0.34*	–0.22*	–
CLASS Pre-K Emotional Support	–	–	–	–
CLASS Pre-K Classroom Organization	–0.36*	–	–0.56*	–
CLASS Pre-K Instructional Support	–	–	–0.33*	–0.51**
CLASS Toddler Emotional and Behavioral Support	–	n.a.	–	–
CLASS Toddler Engaged Support for Learning	–	n.a.	–	–
CIS	–	–	–	–

SOURCE: Authors' analysis of provider sample classroom observations and Delaware Stars administrative data.
NOTES: Difference in average classroom observation score by program characteristic is regression-adjusted from models described in Appendix E. A single asterisk (*) denotes that the difference in the regression-adjusted average score for having the characteristic versus not having the characteristic is significantly different at $p < 0.05$ prior to adjusting for multiple hypothesis testing. A double asterisk (**) indicates the difference remains significant after adjusting for multiple hypothesis testing. – = not statistically significant prior to adjusting for multiple hypothesis testing. n.a. = not applicable.

Supplemental Tables for Chapter 4

Table F.7. Number of Children Completing Assessment in Fall 2014 and Spring 2015

Developmental Outcome (Spring 2015)	Not in Stars	Starting with Stars or Star 2	Star 3	Star 4	Star 5
Youngest cohort (born September 1, 2011, to August 31, 2012)					
PPVT	18	34	16	55	89
WJ–Letter Word Identification	20	34	19	64	94
WJ–Applied Problems	20	35	18	60	93
HTKS	6	8	1	15	22
DECA–Absence of Behavior Problems	4	9	1	17	21
DECA–Total Protective Factors	13	33	14	57	85
Middle cohort (born September 1, 2010, to August 31, 2011)					
PPVT	23	46	30	147	137
WJ–Letter Word Identification	23	50	33	161	148
WJ–Applied Problems	23	50	33	147	126
HTKS	22	48	32	139	121
DECA–Absence of Behavior Problems	20	46	34	145	135
DECA–Total Protective Factors	20	46	34	145	135
Oldest cohort (born September 1, 2009, to August 31, 2010)					
PPVT	21	54	24	153	141
WJ–Letter Word Identification	21	56	26	175	152
WJ–Applied Problems	22	53	25	157	136
HTKS	20	52	23	147	133
DECA–Absence of Behavior Problems	19	45	24	155	144
DECA–Total Protective Factors	19	45	24	155	144
All cohorts					
PPVT	62	134	70	355	367
WJ–Letter Word Identification	64	140	78	400	394
WJ–Applied Problems	65	138	76	364	355
HTKS	48	108	56	301	276
DECA–Absence of Behavior Problems	43	100	59	317	300
DECA–Total Protective Factors	52	124	72	357	364

SOURCES: RAND-administered child assessments and Delaware Stars data.
NOTES: The number of children tested differed by assessment because of Spanish language screener rules, child absences during test administration and make-up periods, and child refusals such that some children took some, but not all, assessments.

Table F.8. Regression-Adjusted Average Child Developmental Assessment Standardized Score by Quartiles of Quality Domain Scores and ERS Score in Delaware Stars Rating

	Quartile of Quality Domain Score or ERS Score			
Developmental Outcome (Spring 2015)	**Lowest**	**Second-Lowest**	**Second-Highest**	**Highest**
Family and Community Partnerships				
PPVT	6.76	6.76	6.65	6.68
WJ–Letter Word Identification	10.60	10.56	10.50	10.54
WJ–Applied Problems	14.93	15.03	15.01	14.98
HTKS	0.82	0.85	0.76	1.00
DECA–Absence of Behavior Problems	5.30	5.46	5.30	5.52
DECA–Total Protective Factors	5.49	5.83	5.93	5.91
Qualifications and Professional Development				
PPVT	6.73	6.76	6.71	6.68
WJ–Letter Word Identification	10.55	10.44[D]	10.49[d]	10.62[B,c]
WJ–Applied Problems	14.87[d]	14.85[D]	15.03	15.10[a,B]
HTKS	0.71	0.89	1.00	0.87
DECA–Absence of Behavior Problems	5.32	5.12	5.43	5.56
DECA–Total Protective Factors	5.79	5.61	5.84	5.88
Management and Administration				
PPVT	6.71	6.67	6.65	6.81
WJ–Letter Word Identification	10.53	10.53	10.48	10.64
WJ–Applied Problems	14.87	15.05	14.97	15.04
HTKS	0.67[b,D]	0.93[a]	0.78	0.99[A]
DECA–Absence of Behavior Problems	5.34	5.32	5.52	5.47
DECA–Total Protective Factors	5.77	5.80	5.87	5.80
Learning Environment and Curriculum				
PPVT	6.76	6.66	6.73	6.71
WJ–Letter Word Identification	10.59	10.44	10.60	10.59
WJ–Applied Problems	15.02	14.90	15.09	15.00
HTKS	0.71	0.82	0.90	0.99
DECA–Absence of Behavior Problems	5.27	5.29	5.38	5.60
DECA–Total Protective Factors	5.73	5.76	5.94	5.81
ERS Total Score				
PPVT	6.68	6.82	6.79	6.77
WJ–Letter Word Identification	10.57	10.53	10.53	10.54
WJ–Applied Problems	14.93	15.14	15.18	15.02
HTKS	0.80	0.87	0.95	0.85
DECA–Absence of Behavior Problems	5.59	5.26	5.49	5.32
DECA–Total Protective Factors	5.78	5.79	6.01	5.72

SOURCES: Authors' analysis of provider sample classroom observations and Delaware Stars administrative data.
NOTES: Sample restricted to programs at Star 3 to Star 5 excluding those rated through an alternative pathway. Average scores by quartile of the quality domain scores are regression-adjusted from models described in Appendix E. Superscript letters denote that the difference in the regression-adjusted average score is significantly different at $p < 0.05$ prior to adjusting for multiple hypothesis testing from [a] lowest quartile, [b] second-lowest quartile, [c] second-highest quartile, and [d] highest quartile. Capitalized superscript letters denote differences that remain statistically significant after adjusting for multiple hypothesis testing.

Table F.9. Regression-Adjusted Difference in Average Child Developmental Assessment Standardized Score for Programs Meeting an Essential Standard Versus Programs Not Meeting the Standard in Delaware Stars Rating

	Essential Standard					
Alternative Program Quality Measure	**Annual Child Developmental Screening**	**Twice-Annual Formative Child Assessment**	**Written Comprehensive Curriculum**	**Daily Activities, Lesson Plans, Individualized Goal Plans**	**Administrator Credentialed**	**Curriculum Coordinator Credential**
PPVT	–	–	–	–	–	–
WJ–Letter Word Identification	–	0.10*	–	–	–	–
WJ–Applied Problems	–	–	–	–	–	–
HTKS	–	–	0.23*	–	0.20*	–
DECA–Absence of Behavior Problems	–	–	–	–	–	–
DECA–Total Protective Factors	0.37*	–	–	–	–	–

SOURCES: Authors' analysis of provider sample classroom observations and Delaware Stars administrative data.
NOTES: Sample restricted to programs at Star 3 to Star 5 excluding those rated through an alternative pathway. Difference in average classroom observation score by meeting or not meeting each essential standard is regression-adjusted from models described in Appendix E. A single asterisk (*) denotes that the difference in the regression-adjusted average score for meeting the standard versus not meeting the standard is significantly different at $p < 0.05$ prior to adjusting for multiple hypothesis testing. A double asterisk (**) indicates that the difference remains significant after adjusting for multiple hypothesis testing. – = not statistically significant prior to adjusting for multiple hypothesis testing.

Table F.10. Regression-Adjusted Average Child Developmental Assessment Standardized Score by Quartiles of Sum of Points for Essential Standards in Delaware Stars Rating

	Quartile of Sum of Points for Essential Standards			
Alternative Program Quality Measure	**Lowest**	**Second-Lowest**	**Second-Highest**	**Highest**
PPVT	–	–	–	–
WJ–Letter Word Identification	–	–	–	–
WJ–Applied Problems	–	–	–	–
HTKS	0.74[D]	0.70[D]	0.82	1.07[A,B]
DECA–Absence of Behavior Problems	–	–	–	–
DECA–Total Protective Factors	–	–	–	–

SOURCES: Authors' analysis of provider sample classroom observations and Delaware Stars administrative data.
NOTES: Sample restricted to programs at Star 3 to Star 5 excluding those rated through an alternative pathway. Average scores by quartile of sum of points for essential standards are regression-adjusted from models described in Appendix E. Superscript letters denote that the difference in the regression-adjusted average score is significantly different at $p < 0.05$ prior to adjusting for multiple hypothesis testing from [a] lowest quartile, [b] second-lowest quartile, [c] second-highest quartile, and [d] highest quartile. Capitalized superscript letters denote differences that remain statistically significant after adjusting for multiple hypothesis testing. – = not statistically significant prior to adjusting for multiple hypothesis testing.

Appendix G. Sensitivity Analyses

This appendix describes sensitivity analyses related to results presented in Chapters 3 and 4.

Sensitivity Analysis for Chapter 3

We performed two sets of sensitivity analyses for results presented in Chapter 3.

Exclusion of Programs Rated Through Alternative Pathway

As noted in Chapter 1, some programs receive their Delaware Stars rating through an alternative pathway that does not require the same degree of quality assessment, either through an ERS assessment or determination of the points-based quality standards. For this reason, we may not expect the same relationship to hold between Delaware Stars ratings and the alternative measures of program quality for programs rated through an alternative pathway compared with those rated through the standard rating process. We therefore estimated the preferred model from Chapter 3 (e.g., Table 3.2) excluding the alternative pathway programs. Results are shown in Table G.1, and the findings presented in Chapter 3 are not affected.

Table G.1. Regression-Adjusted Average Score on Alternative Quality Measures by Delaware Stars Rating Level When Alternative Certification Programs Are Excluded

Alternative Program Quality Measure	Not in Stars	Starting with Stars or Star 2	Star 3	Star 4	Star 5
PQA	3.21	3.13[E]	3.23[E]	3.40	3.58[B,C]
CLASS Pre-K Emotional Support	5.22	5.57	5.46	5.72	5.57
CLASS Pre-K Classroom Organization	4.56	4.62	4.58	5.03	4.82
CLASS Pre-K Instructional Support	1.96[E]	2.31[c]	1.76[b,E]	2.05	2.49[A,C]
CLASS Toddler Emotional and Behavioral Support	4.99	4.85	5.02	5.40	5.57
CLASS Toddler Engaged Support for Learning	2.46[e]	2.51[e]	2.59[e]	2.87	3.27[a,b,c]
CIS	3.24	3.28	3.29	3.33	3.38
Number of programs					
Where PQA observations completed	12	31	19	44	31
Where CLASS Pre-K observations completed	12	32	20	44	31
Where CLASS Toddler observations completed	10	28	19	34	22
Where CIS observations completed	11	32	20	43	31

SOURCES: Authors' analysis of provider sample and Delaware Stars administrative data.
NOTES: There were 32 alternative certification programs in the sample that are excluded from this analysis. When we exclude those 32 programs, the number of assessed children declines from 1,123 to 812. Superscript letters denote that the difference in the regression-adjusted average score is significantly different at $p < 0.05$ prior to adjusting for multiple hypothesis testing from [a] not in Stars, [b] Starting with Stars or Star 2, [c] Star 3, [d] Star 4, and [e] Star 5. Capitalized superscript letters denote differences that remain statistically significant after adjusting for multiple hypothesis testing.

Alternative Approach to Aggregating Across Classrooms

With the exception of the CIS, which is a teacher-level measure, all of our alternative quality measures are initially measured at the classroom level. We observed up to three classrooms per program. In Chapter 3, we reported the associations between average program quality scores (e.g., the average of up to three classroom observations) and the program's Delaware Stars rating. But the Delaware Stars rating system uses the minimum ERS score across the assessed classrooms, not the average score. If there is little variation in measured quality across classrooms in the same program, our results should not be sensitive to whether we use the average classroom score to represent the level of quality for the provider as a whole or whether we use the minimum classroom score.

However, as shown in Table G.2, there is considerable variation across classrooms in quality as measured by the PQA, the CLASS subscales, and the CIS.[39] For example, we assessed more than one classroom using the PQA for 100 providers. The average difference between the lowest- and highest-scoring classroom was 0.6 scale points on the five-point PQA scale, the equivalent of about one scale point on the seven-point CLASS scale or 0.5 scale points on the four-point CIS scale. The largest gap between the minimum- and maximum-scoring classroom on the PQA was 2.0 scale points, and the smallest gap was 0.01 scale points. The average difference between the lowest- and highest-scoring classrooms on the CLASS subscales ranged from 0.9 to 1.3 scale points, with gaps as large as 4 scale points. For the CIS, the average gap was 0.5 scale points, with gaps as large as 2 scale points. Thus, on each measure, the average gap between the lowest- and highest-scoring classroom was the equivalent of one scale point on a

Table G.2. Variation in Quality Scores Across Classrooms in the Same Program for Programs Where Multiple Classrooms Were Assessed

Alternative Program Quality Measure	Number of Programs with Two or More Classrooms Assessed	Average Difference in Scores Across Classrooms	Largest Difference in Scores Across Classrooms	Smallest Difference in Scores Across Classrooms
PQA	100	0.60	2.02	0.01
CLASS Pre-K Emotional Support	98	0.95	3.66	0.00
CLASS Pre-K Classroom Organization	98	1.29	4.17	0.00
CLASS Pre-K Instructional Support	98	0.88	3.25	0.00
CLASS Toddler Emotional and Behavioral Support	4	1.07	2.10	0.13
CLASS Toddler Engaged Support for Learning	4	1.13	1.55	0.87
CIS	115	0.53	2.02	0.01

SOURCES: Authors' analysis of provider sample and Delaware Stars administrative data.

[39] This finding is consistent with analyses in other jurisdictions of the extent of variation in quality across classrooms in the same program (see Karoly, Zellman, and Perlman, 2013).

seven-point scale. With this degree of difference across classrooms in the same program, it is important to assess the sensitivity of our analyses to the use of the center average score versus the minimum score.

Table G.3 reports the results when using the lowest-scoring classroom for each provider in the models we report in Chapter 3 (see Table 3.1). The results are substantively the same, except there are fewer statistically significant differences after adjusting for multiple hypothesis testing.

We also specified a regression where each classroom served as the unit of analyses, rather than the provider, and we clustered the standard errors to account for the clustering of classrooms within programs. Table G.4 reports the results. The same instruments bear approximately the same associations as are shown in Chapter 3.

Sensitivity Analyses for Chapter 4

We estimated a number of alternative model specifications for the results presented in Chapter 4.

Specification of the Outcome Measure

Other validation studies have examined the relationship between child outcomes and program QRIS ratings using gain scores, rather than the specification we employ. Other studies have used hierarchical linear modeling as well. We estimated these alternative specifications, controlling for the same set of covariates in our preferred model.

Table G.3. Regression-Adjusted Average Score on Alternative Quality Measures by Delaware Stars Rating Level When Using Minimum Classroom Value

Alternative Program Quality Measure	Not in Stars	Starting with Stars or Star 2	Star 3	Star 4	Star 5
PQA	3.04	2.88[E]	3.09[e]	3.17[e]	3.38[B,c,d]
CLASS Pre-K Emotional Support	4.83	5.28	5.33	5.41	5.41
CLASS Pre-K Classroom Organization	4.03	4.20	4.36	4.76	4.71
CLASS Pre-K Instructional Support	1.94	1.96	1.68[E]	1.84[e]	2.26[C,d]
CLASS Toddler Emotional and Behavioral Support	5.00	4.88	4.97	5.38	5.56
CLASS Toddler Engaged Support for Learning	2.56	2.52[e]	2.53[e]	2.81	3.21[b,c]
CIS	3.01	3.05	3.10	3.12	3.24
Number of programs					
Where PQA observations completed	13	31	19	57	42
Where CLASS Pre-K observations completed	13	32	20	60	44
Where CLASS Toddler observations completed	11	28	19	34	27
Where CIS observations completed	12	32	20	59	44

SOURCES: Authors' analysis of provider sample and Delaware Stars administrative data.
NOTES: Superscript letters denote that the difference in the regression-adjusted average score is significantly different at $p < 0.05$ prior to adjusting for multiple hypothesis testing from [a] not in Stars, [b] Starting with Stars or Star 2, [c] Star 3, [d] Star 4, and [e] Star 5. Capitalized superscript letters denote differences that remain statistically significant after adjusting for multiple hypothesis testing.

Table G.4. Regression-Adjusted Average Score on Alternative Quality Measures by Delaware Stars Rating Level When Using Classroom-Level Observations

Alternative Program Quality Measure	Not in Stars	Starting with Stars or Star 2	Star 3	Star 4	Star 5
PQA	3.10[E]	3.04[E]	3.16[E]	3.26[E]	3.52[A,B,C,D]
CLASS Pre-K Emotional Support	5.02	5.40	5.17	5.46	5.45
CLASS Pre-K Classroom Organization	4.21	4.37	4.28	4.75	4.63
CLASS Pre-K Instructional Support	2.05[E]	2.40[C]	1.81[B,E]	2.10[E]	2.57[A,C,D]
CLASS Toddler Emotional and Behavioral Support	4.87	4.75	4.87	5.26	5.45
CLASS Toddler Engaged Support for Learning	2.48	2.43[e]	2.47[e]	2.73	3.11[b,c]
CIS	3.16	3.24	3.22	3.29	3.38
Number of classrooms					
Where PQA observations completed	23	59	35	108	82
Where CLASS Pre-K observations completed	19	48	29	107	73
Where CLASS Toddler observations completed	11	28	21	35	28
Where CIS observations completed	24	65	38	120	92

SOURCES: Authors' analysis of provider sample and Delaware Stars administrative data.
NOTES: Superscript letters denote that the difference in the regression-adjusted average score is significantly different at $p < 0.05$ prior to adjusting for multiple hypothesis testing from [a] not in Stars, [b] Starting with Stars or Star 2, [c] Star 3, [d] Star 4, and [e] Star 5. Capitalized superscript letters denote differences that remain statistically significant after adjusting for multiple hypothesis testing.

For our first sensitivity analyses, we used gain scores (i.e., spring 2015 score minus fall 2014) as the outcome, rather than the spring 2015 score (controlling for the fall 2014 score). The gain score model requires more assumptions about rate of growth than does the model we prefer. With the gain score models, we accounted for within-provider clustering of outcomes by adjusting the standard errors of estimates using Taylor series linearization, as we did for our main model specification. The formal equation describing these models is

$$D_{ijk} = \alpha_j X_j + \beta_j P_j + \gamma_i C_i + \delta_k N_k + \varepsilon_{ijk},$$

where D_{ijk} is the gain score (spring 2015 score minus fall 2014 score) for child i, attending program j, and residing in neighborhood k; X_j is the main predictor of interest, usually a program-level variable, but sometimes a classroom-level variable; P_j is the vector of program-level covariates, C_i is the vector of child- and family-level covariates (excluding fall 2014 score); N_k is the vector of child neighborhood-level covariates; and ε_{ijk} are child-level errors.

We also specified hierarchical linear models, which controlled for clustering of children's outcomes within provider by including a random intercept for the provider. As for the main model specification, the outcomes are the spring 2015 scores, and fall 2015 scores are included as covariates. The formal model specification is

$$Y_{ijk} = \alpha_j X_j + \beta_j P_j + \gamma_i C_i + \delta_k N_k + \zeta_j + \varepsilon_{ijk},$$

where Y_{ijk} is the spring 2015 score for child i, attending program j, and residing in neighborhood k; ζ_j are program-level errors (also known as the random program-level effect); X_j is the main predictor of interest, usually a program-level variable, but sometimes a classroom-level variable;

P_j is the vector of program-level covariates, C_i is the vector of child- and family-level covariates; N_k is the vector of child neighborhood-level covariates; and ε_{ijk} are child-level errors. Results from hierarchical models did not differ substantively from the results presented in the report. We compared results from these two alternative model specifications to the main model primarily by comparing the *p*-values of joint tests of the predictor of primary interest. There were no substantial differences in our findings based on these alternative specifications.

Timing of Delaware Stars Rating

In our preferred specification of the Delaware Stars rating status, we use their status as of August 2015 to categorize programs as not in Stars, Starting with Stars/Star 2, Star 3, Star 4, or Star 5. We tested whether Star 4 and Star 5 programs that had had those ratings since May 2014 were different from programs that had attained those high ratings between May 2014 and August 2015. Specifically, we compared star ratings in May 2014 and August 2015 and recategorized programs at Star 4 (Star 5) in August 2015 as having either moved to that level since May 2014 or being rated at Star 4 (Star 5) at both time points. We found no difference between stable and recent Star 4 programs or between stable and recent Star 5 programs.

Alternative Functional Form for Delaware Stars Domains

As described in Appendix E, to examine the relationship between quality as measured by the four Delaware Stars domains and child developmental outcomes, our preferred methodology grouped programs into quartiles based on the number of points the programs obtained in a given domain, where quartiles accounted for the differences in points-based standards for centers versus FCCs and before and after the July 2014 revisions. As an alternate specification, we estimated the same models of child developmental outcomes exclusively for center-based programs, since they had the smallest change in number of points that could be obtained per domain in the pre– versus post–July 2014 Delaware Stars standards. The models included a continuous measure of the points obtained in the domain, rather than the quartile specification. With this alternative specification, we find substantively the same results as presented in Chapter 4: Namely, the number of points obtained in a given domain generally does not predict meaningful differences in children's outcomes. The finding in Table F.8 of a relationship between Qualifications and Professional Development and the WJ–Applied Problems assessment is replicated in the continuous linear model, but not the finding for WJ–Letter Word Identification.

Approach to Aggregating Observational Measures

We also performed sensitivity analyses on the method for linking classroom-level (PQA, CLASS) and teacher-level (CIS) observation measures to child-level data. In our preferred model specification, when there were multiple classrooms with observation measures (PQA, CLASS),

we took the mean of classroom observations within a program and then merged those means to the child-level data by program. For the CIS, we first took the average of the CIS scores within the classroom and then constructed the overall program mean using the classroom-level means.

We examined the sensitivity of our findings to two alternative methods of combining observational and child-assessment information. In the first, the minimum score across classrooms (or staff) was taken for each program (or classroom), then merged to the child-level data. These minimum values are the lowest quality care that a program was observed to provide. In the second method, we merged CLASS and PQA classroom-level observations to child-level data by classroom. The mean of the CIS score was taken within each classroom, then merged to the child-level data. This classroom-level merge allowed modeling of child assessment scores on observational data from their own classroom. Of the 1,123 children assessed in both fall 2014 and spring 2015, 872 children (82 percent) were in a classroom with at least one observational score. Again, both alternative specifications produced findings that were similar to our preferred model reported in Chapter 4.

Exclusion of Subsets of Programs and Children

We assessed whether results differed when excluding subsets of programs and children. Namely, in Table G.5, we present results when excluding the 32 alternative certification programs in the study sample, as we did for the Chapter 3 analysis above. Tables G.6 and G.7 present main results when excluding the youngest cohort of children and including only the oldest cohort, samples that correspond to those used in other QRIS validation studies that focus on children one or two years before kindergarten entry. In another analysis, we excluded children for whom English was not the language spoken at home, as reported in the parent survey (see Table G.8). Again, in each case, our findings are not substantially affected by these alternative models.

Table G.5. Regression-Adjusted Average Child Developmental Assessment Standardized Score by Delaware Stars Rating Level When Alternative Certification Programs Are Excluded

Developmental Outcome (Spring 2015)	Not in Stars	Starting with Stars or Star 2	Star 3	Star 4	Star 5
PPVT	6.49	6.50	6.50	6.48	6.45
WJ–Letter Word Identification	10.56	10.71	10.54	10.61	10.68
WJ–Applied Problems	15.17	15.20	14.94	15.05	15.08
HTKS	0.83[b]	0.60[a,c,E]	0.90[b]	0.79[E]	1.05[B,D]
DECA–Absence of Behavior Problems	5.52	5.15	5.09	5.18	5.39
DECA–Total Protective Factors	6.17	5.85	5.83	5.68	5.82
Number of programs where children were assessed	10	27	14	32	25

SOURCES: Authors' analysis of child sample assessments and Delaware Stars administrative data.
NOTES: There were 32 alternative certification programs in the sample that are excluded from this analysis. When we exclude those 32 programs, the number of assessed children declines from 1,123 to 812. Superscript letters denote that the difference in the regression-adjusted average score is significantly different at $p < 0.05$ prior to adjusting for multiple hypothesis testing from [a] not in Stars, [b] Starting with Stars or Star 2, [c] Star 3, [d] Star 4, and [e] Star 5. Capitalized superscript letters denote differences that remain statistically significant after adjusting for multiple hypothesis testing.

Table G.6. Regression-Adjusted Average Child Developmental Assessment Standardized Score by Delaware Stars Rating Level When the Youngest Age Cohort Is Excluded

Developmental Outcome (Spring 2015)	Not in Stars	Starting with Stars or Star 2	Star 3	Star 4	Star 5
PPVT	7.39[c,D,E]	7.22	7.16[a]	7.13[A]	7.17[A]
WJ–Letter Word Identification	11.36	11.47	11.37	11.40	11.42
WJ–Applied Problems	18.55	18.31	18.24	18.24	18.28
HTKS	0.93	0.73[E]	0.98	0.88[e]	1.07[B,d]
DECA–Absence of Behavior Problems	5.63	5.38	5.23	5.37	5.46
DECA–Total Protective Factors	6.47	6.28	6.10	5.99	6.12
Number of children assessed	46	110	61	351	312

SOURCES: Authors' analysis of child sample assessments and Delaware Stars administrative data.
NOTES: The youngest cohort refers to children born September 2, 2011, to August 31, 2012. When this cohort is excluded, the sample declines from 1,123 to 880. For each rating level, we report the number of children with at least one assessment. Superscript letters denote that the difference in the regression-adjusted average score is significantly different at $p < 0.05$ prior to adjusting for multiple hypothesis testing from [a] not in Stars, [b] Starting with Stars or Star 2, [c] Star 3, [d] Star 4, and [e] Star 5. Capitalized superscript letters denote differences that remain statistically significant after adjusting for multiple hypothesis testing.

Table G.7. Regression-Adjusted Average Child Developmental Assessment Standardized Score by Delaware Stars Rating Level When Only the Oldest Age Cohort Is Included

Developmental Outcome (Spring 2015)	Not in Stars	Starting with Stars or Star 2	Star 3	Star 4	Star 5
PPVT	8.82	8.55	8.60	8.49	8.60
WJ–Letter Word Identification	11.63	11.64	11.60	11.63	11.61
WJ–Applied Problems	20.77	20.48	20.28	20.50	20.60
HTKS	1.54	0.98	1.18	1.28	1.46
DECA–Absence of Behavior Problems	5.69	5.33	5.09	5.44	5.50
DECA–Total Protective Factors	6.89	6.35	6.06	5.94	6.17
Number of children assessed	22	58	26	185	160

SOURCES: Authors' analysis of child sample assessments and Delaware Stars administrative data.
NOTES: The oldest cohort refers to children born September 1, 2009, to August 31, 2010. When only this cohort is included, the sample declines from 1,123 to 451. For each rating level, we report the number of children with at least one assessment. Superscript letters denote that the difference in the regression-adjusted average score is significantly different at $p < 0.05$ prior to adjusting for multiple hypothesis testing from [a] not in Stars, [b] Starting with Stars or Star 2, [c] Star 3, [d] Star 4, and [e] Star 5. Capitalized superscript letters denote differences that remain statistically significant after adjusting for multiple hypothesis testing.

Relationship Between Child Development and Alternative Measures of Program Quality

Tables G.9 to G.15 show the results from estimating models to examine the relationship between each of the child development measures and the alternative measures of program quality (PQA, CLASS, and CIS). For these analyses, each quality measure was categorized into five or six approximately equal-sized groups based on program-level average scores to allow for possible nonlinearities in the relationship between program quality and child developmental outcomes. These models also included controls for group size and the adult-child ratio in the observed classrooms. In addition, in the models with the PQA as the predictor, we included an indicator

variable for whether the assessment was conducted with the PQA pre-K instrument. These results demonstrate that, with some exceptions, there is not a strong positive relationship between the alternative measures of program quality and child outcomes.

Table G.8. Regression-Adjusted Average Child Developmental Assessment Standardized Score by Delaware Stars Rating Level When Only Children Who Speak English at Home Are Included

Developmental Outcome (Spring 2015)	Not in Stars	Starting with Stars or Star 2	Star 3	Star 4	Star 5
PPVT	6.79	6.84	6.81	6.78	6.76
WJ–Letter Word Identification	11.00	11.26	11.01	11.10	11.14
WJ–Applied Problems	15.32	15.44	15.10	15.25	15.25
HTKS	0.86	0.74	0.91	0.84	1.03
DECA–Absence of Behavior Problems	5.67	5.29	5.32	5.39	5.59
DECA–Total Protective Factors	6.30	5.85	5.95	5.79	5.92
Number of children assessed	62	130	78	337	345

SOURCES: Authors' analysis of child sample assessments and Delaware Stars administrative data.
NOTES: When only children who speak English at home are included, the sample declines from 1,123 to 952. For each rating level, we report the number of children with at least one assessment. Superscript letters denote that the difference in the regression-adjusted average score is significantly different at $p < 0.05$ prior to adjusting for multiple hypothesis testing from [a] not in Stars, [b] Starting with Stars or Star 2, [c] Star 3, [d] Star 4, and [e] Star 5. Capitalized superscript letters denote differences that remain statistically significant after adjusting for multiple hypothesis testing.

Table G.9. Regression-Adjusted Average Child Developmental Assessment Standardized Score by PQA Score

Developmental Outcome (Spring 2015)	Lowest Score [1–2.5)	Second-Lowest Score [2.5–3)	Third-Lowest Score [3–3.5)	Second-Highest Score [3.5–4)	Highest Score [4–5]
PPVT	6.31	6.33	6.36	6.39	6.33
WJ–Letter Word Identification	10.66	10.62	10.63	10.71	10.73
WJ–Applied Problems	15.13	14.98	15.02	14.97	15.05
HTKS	0.93	0.85	0.91	0.95	0.81
DECA–Absence of Behavior Problems	5.56	5.35	5.44	5.37	5.72
DECA–Total Protective Factors	5.96	5.84	5.91	5.74	5.89

SOURCES: RAND-administered child assessments and RAND classroom observations.
NOTES: Superscript letters denote that the difference in the regression-adjusted average scores is significantly different prior to adjustment for multiple hypothesis testing from [a] lowest of five score categories, [b] second-lowest of five score categories, [c] third-lowest of five score categories, [d] second-highest of five score categories, and [e] highest of five score categories. Capitalized superscript letters denote differences that remain statistically significant after adjusting for multiple hypothesis testing. When expressing score ranges, a square bracket denotes the inclusion of the starting or ending value, and a round bracket denotes noninclusion of the value.

Table G.10. Regression-Adjusted Average Child Developmental Assessment Standardized Score by CLASS Pre-K Emotional Support Score

Developmental Outcome (Spring 2015)	Lowest Score [1–4)	Second-Lowest Score [4–4.5)	Third-Lowest Score [4.5–5)	Third-Highest Score [5–5.5)	Second-Highest Score [5.5–6)	Highest Score [6–7]
PPVT	6.31	6.51	6.38	6.33	6.38	6.32
WJ–Letter Word Identification	10.75	10.59	10.65	10.70	10.68	10.68
WJ–Applied Problems	15.13	15.17	15.12	15.00	15.11	15.08
HTKS	0.74	1.21	0.86	0.93	0.91	0.88
DECA–Absence of Behavior Problems	5.06	5.36	5.26	5.26	5.42	5.57
DECA–Total Protective Factors	5.83	5.96	5.81	5.89	5.92	5.93

SOURCES: RAND-administered child assessments and RAND classroom observations.
NOTES: Superscript letters denote that the difference in the regression-adjusted average scores is significantly different prior to adjustment for multiple hypothesis testing from [a] lowest of six score categories, [b] second-lowest of six score categories, [c] third-lowest of six score categories, [d] third-highest of six score categories, [e] second-highest of six score categories, and [f] highest of six score categories. Capitalized superscript letters denote differences that remain statistically significant after adjusting for multiple hypothesis testing. When expressing score ranges, a square bracket denotes the inclusion of the starting or ending value, and a round bracket denotes noninclusion of the value.

Table G.11. Regression-Adjusted Average Child Developmental Assessment Standardized Score by CLASS Pre-K Classroom Organization Score

Developmental Outcome (Spring 2015)	Lowest Score [1–2.5)	Second-Lowest Score [2.5–4)	Third-Lowest Score [4–4.5)	Third-Highest Score [4.5–5)	Second-Highest Score [5–5.5)	Highest Score [5.5–7]
PPVT	6.36	6.34	6.36	6.35	6.44	6.33
WJ–Letter Word Identification	10.86	10.61	10.66	10.69	10.69	10.70
WJ–Applied Problems	15.10	15.03	15.16	15.03	15.11	15.13
HTKS	1.01	0.89	0.99	0.84 [e]	1.03 [d,F]	0.81[E]
DECA–Absence of Behavior Problems	5.36	5.22	5.21	5.49	5.38	5.54
DECA–Total Protective Factors	6.29[b]	5.72[a,d]	5.78	6.02[b]	5.98	5.91

SOURCES: RAND-administered child assessments and RAND classroom observations.
NOTES: Superscript letters denote that the difference in the regression-adjusted average scores is significantly different prior to adjustment for multiple hypothesis testing from [a] lowest of six score categories, [b] second-lowest of six score categories, [c] third-lowest of six score categories, [d] third-highest of six score categories, [e] second-highest of six score categories, and [f] highest of six score categories. Capitalized superscript letters denote differences that remain statistically significant after adjusting for multiple hypothesis testing. When expressing score ranges, a square bracket denotes the inclusion of the starting or ending value, and a round bracket denotes noninclusion of the value.

Table G.12. Regression-Adjusted Average Child Developmental Assessment Standardized Score by CLASS Pre-K Instructional Support Score

Developmental Outcome (Spring 2015)	Lowest Score [1–1.5)	Second-Lowest Score [1.5–2)	Third-Lowest Score [2–2.5)	Second-Highest Score [2.5–3)	Highest Score [3–7]
PPVT	6.26	6.37	6.39	6.36	6.35
WJ–Letter Word Identification	10.69	10.62	10.64	10.76	10.73
WJ–Applied Problems	14.99	15.08	15.11	15.16	15.07
HTKS	0.85	0.92	0.92	0.92	0.88
DECA–Absence of Behavior Problems	5.39	5.28	5.46	5.51	5.26
DECA–Total Protective Factors	5.72[d]	5.90[d]	5.98[E]	6.11[a, b, E]	5.64[C, D]

SOURCES: RAND-administered child assessments and RAND classroom observations.
NOTES: Superscript letters denote that the difference in the regression-adjusted average scores is significantly different prior to adjustment for multiple hypothesis testing from [a] lowest of five score categories, [b] second-lowest of five score categories, [c] third-lowest of five score categories, [d] second-highest of five score categories, and [e] highest of five score categories. Capitalized superscript letters denote differences that remain statistically significant after adjusting for multiple hypothesis testing. When expressing score ranges, a square bracket denotes the inclusion of the starting or ending value, and a round bracket denotes noninclusion of the value.

Table G.13. Regression-Adjusted Average Child Developmental Assessment Standardized Score by CLASS Toddler Emotional and Behavioral Support Score

Developmental Outcome (Spring 2015)	Lowest Score [1–4)	Second-Lowest Score [4–4.5)	Third-Lowest Score [4.5–5)	Third-Highest Score [5–5.5)	Second-Highest Score [5.5–6)	Highest Score [6–7]
PPVT	6.29[b]	6.57[a, D]	6.37	6.31[B, e]	6.44[d]	6.43
WJ–Letter Word Identification	10.36[B,D,f]	10.10[A, e]	10.25	10.18[A]	10.26[b]	10.20[a]
WJ–Applied Problems	14.41[c]	14.59	14.69[a, D]	14.44[C,e,F]	14.65[d]	14.66[D]
HTKS	0.71	0.94	0.78	0.99	0.94	0.81
DECA–Absence of Behavior Problems	5.52	5.05	5.19	5.45	5.34	5.27
DECA–Total Protective Factors	5.76	5.47	5.62	5.76	5.74	5.72

SOURCES: RAND-administered child assessments and RAND classroom observations.
NOTES: Superscript letters denote that the difference in the regression-adjusted average scores is significantly different prior to adjustment for multiple hypothesis testing from [a] lowest of six score categories, [b] second-lowest of six score categories, [c] third-lowest of six score categories, [d] third-highest of six score categories, [e] second-highest of six score categories, and [f] highest of six score categories. Capitalized superscript letters denote differences that remain statistically significant after adjusting for multiple hypothesis testing. When expressing score ranges, a square bracket denotes the inclusion of the starting or ending value, and a round bracket denotes noninclusion of the value.

Table G.14. Regression-Adjusted Average Child Developmental Assessment Standardized Score by CLASS Toddler Engaged Support for Learning Score

Developmental Outcome (Spring 2015)	Lowest Score [1–1.5)	Second-Lowest Score [1.5–2)	Third-Lowest Score [2–2.5)	Second-Highest Score [2.5–3)	Highest Score [3–7]
PPVT	6.22	6.34	6.42	6.35	6.42
WJ–Letter Word Identification	10.18	10.22	10.22	10.27	10.22
WJ–Applied Problems	14.39	14.44	14.57	14.56	14.61
HTKS	0.80	0.71	0.84	0.92	0.94
DECA–Absence of Behavior Problems	4.76	5.47	5.40	5.53	5.21
DECA–Total Protective Factors	5.78	5.67	5.68	5.82	5.68

SOURCES: RAND-administered child assessments and RAND classroom observations.
NOTES: Superscript letters denote that the difference in the regression-adjusted average scores is significantly different prior to adjustment for multiple hypothesis testing from [a] lowest of five score categories, [b] second-lowest of five score categories, [c] third-lowest of five score categories, [d] second-highest of five score categories, and [e] highest of five score categories. Capitalized superscript letters denote differences that remain statistically significant after adjusting for multiple hypothesis testing. When expressing score ranges, a square bracket denotes the inclusion of the starting or ending value, and a round bracket denotes noninclusion of the value.

Table G.15. Regression-Adjusted Average Child Developmental Assessment Standardized Score by CIS Score

Developmental Outcome (Spring 2015)	Lowest Score [1–2.5)	Second-Lowest Score [2.5–3)	Third-Lowest Score [3–3.25)	Second-Highest Score [3.25–3.5)	Highest Score [3.5–4]
PPVT	6.13	6.35	6.31	6.40	6.37
WJ–Letter Word Identification	10.66	10.65	10.69	10.67	10.67
WJ–Applied Problems	14.66	15.14	15.08	15.05	15.12
HTKS	0.96	0.95	0.89	0.95	0.87
DECA–Absence of Behavior Problems	5.28	5.28	5.28	5.36	5.45
DECA–Total Protective Factors	5.79	5.79	5.81	5.83	5.99

SOURCES: RAND-administered child assessments and RAND classroom observations.
NOTES: Superscript letters denote that the difference in the regression-adjusted average scores is significantly different prior to adjustment for multiple hypothesis testing from [a] lowest of five score categories, [b] second-lowest of five score categories, [c] third-lowest of five score categories, [d] second-highest of five score categories, and [e] highest of five score categories. Capitalized superscript letters denote differences that remain statistically significant after adjusting for multiple hypothesis testing. When expressing score ranges, a square bracket denotes the inclusion of the starting or ending value, and a round bracket denotes noninclusion of the value.

References

American Association for Public Opinion Research. (2015). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. 8th edition. AAPOR. As of July 1, 2016: https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions2015_8theditionwithchanges_April2015_logo.pdf

Arnett, J. (1989). Caregivers in day-care centers: Does training matter? *Journal of Applied Developmental Psychology* 10(4): 541–552.

Auger, A., G. Farkas, G. J. Duncan, M. R. Burchinal, G. J. Duncan, and D. L. Vandell. (2014). Preschool center care quality effects on academic achievement: An instrumental variables analysis. *Developmental Psychology* 50(12): 2559.

Auger, A., L. A. Karoly, and H. L. Schwartz. (2015). *Evaluation of Delaware Stars for Early Success: Year 2 Report*, RR-1026-DOEL. Santa Monica, CA: RAND Corporation. As of July 1, 2016: http://www.rand.org/pubs/research_reports/RR1026.html

Benjamini, Y., and Y. Hochberg. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)* 57(1): 289–300.

BUILD Initiative. (2016). *QRIS Compendium*. Website. As of July 1, 2016: http://qriscompendium.org/

Burchinal, M. R. (2010). *Differentiating Among Measures of Quality: Key Characteristics and Their Coverage in Existing Measures*. OPRE Research-to-Policy, Research-to-Practice Brief OPRE 2011-10b. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. As of July 1, 2016: https://www.acf.hhs.gov/sites/default/files/opre/differ_measures.pdf

Burchinal, M. R., K. Kainz, and Y. Cai. (2011). How well do our measures of quality predict child outcomes? A meta-analysis and coordinated analysis of data from large scale studies of early childhood settings. In M. Zaslow, I. Martinez-Beck, K. Tout, and T. Halle (Eds.), *Quality Measurement in Early Childhood Settings* (pp. 11–31). Baltimore, MD: Brookes.

Burchinal, M. R., L. Tarullo, and M. Zaslow, (2016). *Best practices in creating and adapting Quality Rating and Improvement System (QRIS) rating scales*. OPRE Research Brief #2016-25. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. As of July 1, 2016: http://www.acf.hhs.gov/sites/default/files/opre/cceepra_qris_531_508compliant_66_b508.pdf

Cameron, C. E. and M. M. McClelland. (2011). *Head-Toes-Knees-Shoulders (HTKS)—Extended*. Protocol obtained with permission from the authors.

———. (2011). *Cabeza-Dedos del pie-Rodillas-Hombros (HTKS)—Extendido*. Protocol obtained with permission from the authors.

Child Care Aware. (2013). *We Can Do Better: Child Care Aware of America's Ranking of State Child Care Regulations and Oversight*. As of July 1, 2016: http://usa.childcareaware.org/advocacy-public-policy/resources/reports-and-research/we-can-do-better-2013-update/

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.

Cohen, J. (1992). A power primer. *Psychological Bulletin* 112(1): 155–159.

Delaware Institute for Excellence in Early Childhood. (undated). Delaware Stars for Early Success. Website. As of July 1, 2016: http://www.delawarestars.udel.edu/

———. (2014a). Stars enhancements 2014. Newark, DE: Author. As of July 1, 2016: http://www.delawarestars.udel.edu/wp-content/uploads/2014/07/2014.2.27Final-Stars-Enhancements-2014-handout.pdf

———. (2014b). *Stars Enhancements 2014: Continuous Quality Improvement for Early Success*. PowerPoint presentation. Newark, DE: Author. As of July 1, 2016: http://www.delawarestars.udel.edu/wp-content/uploads/2014/07/2014.2.26-Enhancements-PP-Generic-Fianl.pdf

———. (2016). *Delaware Stars for Early Success Monthly Report for May 2016*. Newark, DE: Author.

Duncan, S., and E. DeAvila, E. (2000). *Pre-Language Assessment Scales [PreLAS 2000]*. Monterey, CA: CTB-McGraw Hill.

Dunn, D. M., and L. M. Dunn. (2007). *Peabody Picture Vocabulary Test* (4th ed.). Minneapolis, MN: Pearson Assessments.

Dunn, L. M., D. E. Lugo, E. R. Padilla, and L. R. Dunn. (1986). *Test de Vocabulario en Imagenes Peabody.* Circle Pines, MN: American Guidance Service.

Elicker, J., C. C. Langill, K. Ruprecht, J. Lewsader, and T. Anderson. (2011). *Evaluation of Paths to QUALITY, Indiana's Child Care Quality Rating and Improvement System: Final Report*. West Lafayette, IN: Purdue University. As of July 1, 2016: http://docs.lib.purdue.edu/cffpub/45/

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

Graubard, B. I., and E. L. Korn. (1999). Predictive margins with survey data. *Biometrics,* 55(2): 652–659.

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

Harms, T., D. Cryer, and R. M. Clifford. (2006). *Infant/Toddler Environment Rating Scale–Revised Edition (ITERS-R), Updated*. New York: Teachers College Press.

———. (2007). *Family Child Care Environment Rating Scale–Revised Edition (FCCERS-R)*. New York: Teachers College Press.

High/Scope Educational Research Foundation. (2003). *Preschool Program Quality Assessment, 2nd Edition (PQA)Form A: Classroom Items.* Ypsilanti, MI: HighScope Press.

———. (2009). *Family Child Care Program Quality Assessment, (PQA) Assessment Form.* Ypsilanti, MI: HighScope Press.

Hohmann, M., S. Lockhart, and J. Montie. (2013). *Infant-Toddler Program Quality Assessment (PQA) Form A: Observation Items.* Ypsilanti, MI: HighScope Press.

Karoly, L. A. (2012). *A Golden Opportunity: Advancing California's Early Care and Education Professional Development System*, MG-1188-PF/BECF. Santa Monica, CA: RAND Corporation. As of July 1, 2016: http://www.rand.org/pubs/monographs/MG1188.html

———. (2014). Validation studies for early learning and care quality rating and improvement systems: A review of the literature. WR-1051-DOEL. Santa Monica, CA: RAND Corporation. As of July 1, 2016: http://www.rand.org/pubs/working_papers/WR1051.html

Karoly, L. A., G. L. Zellman, and M. Perlman. (2013). Understanding variation in classroom quality within early childhood centers: Evidence from Colorado's quality rating and improvement system. *Early Childhood Research Quarterly* 28(4): 645–657.

La Paro, K. M., B. K. Hamre, and R. C. Pianta. (2012). *Classroom Assessment Scoring System (CLASS) Manual, Toddler*. Baltimore, MD: Paul H. Brookes Publishing Company.

Lahti, M., C. Cobo-Lewis, A. Dean, S. Rawlings, E. Sawyer, and B. Zollitsch. (2011). *Maine's Quality for Me: Child Care Quality Rating and Improvement System (QRIS): Final Evaluation Report.* Maine: Department of Health and Human Services. As of July 1, 2016: http://muskie.usm.maine.edu/maineroads/pdfs/QRISEVALRPRT_FINAL.pdf

Lahti, M., J. Elicker, G. Zellman, and R. Fiene. (2015). Approaches to validating child care quality rating and improvement systems (QRIS): Results from two states with similar QRIS type designs. *Early Childhood Research Quarterly* 30: 280–290.

LeBuffe, P., V. Shapiro, and J. Naglieri. (2009). *Devereux Student Strengths Assessment (DESSA)*.Villanova, PA.: Devereux Center for Resilient Children.

LeBuffe, P. A., and J. A. Naglieri. (2012). *The Devereux Early Childhood Assessment for Preschoolers, Second Edition (DECA-P2) Assessment, Technical Manual, and User's Guide*. Lewisville, NC: Kaplan.

Mackrain, M., P. A. LeBuffe, and G. Powell. (2007). *The Devereux Early Childhood Assessment for Infants and Toddlers (DECA-I/T) Assessment, Technical Manual, and User's Guide*. Lewisville, NC: Kaplan.

Magnuson, K., and Y. Lin (2015). *Validation of the QRIS YoungStar Rating Scale, Report 1: Wisconsin Early Child Care Study Findings on the Validity of YoungStar Ratings for Observed Classroom Quality*. Madison, WI: School of Social Work and Institute for Research on Poverty. As of July 1, 2016: http://dcf.wisconsin.gov/youngstar/pdf/ys-validation-study/weccs_full_report.pdf

———. *Validation of the QRIS YoungStar Rating Scale, Report 2: Wisconsin Early Child Care Study Findings on the Validity of YoungStar Ratings for Children's School Readiness*. Madison, WI: School of Social Work and Institute for Research on Poverty. As of July 1, 2016: http://dcf.wisconsin.gov/youngstar/pdf/FINAL-Part-2-Validation-Study-Full-Report-2016.pdf

Mashburn, A. J., R. C. Pianta, B. K. Hamre, J. T. Downer, O. A. Barbarin, D. Bryant, M. Burchinal, D. M. Early, and C. Howes. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development*, 79(3): 732–749.

McClelland, M. M., C. E. Cameron, R. Duncan, R. P. Bowles, A. C. Acock, A. Miao, and M. E. Pratt. (2014). Predictors of early growth in academic achievement: The head-toes-knees-shoulders task. *Frontiers in Psychology*, 5: 599.

Mead, S., and A. L. Mitchel. (2016). *Moneyball for Head Start: Using Data, Evidence, and Evaluation to Improve Outcomes for Children and Families*. Washington, DC: Results for America. As of July 1, 2016: http://results4america.org/wp-content/uploads/2016/01/RFA_Headstart_Moneyball_i.pdf

National Center for Early Childhood Quality Assurance. (2015). *Trends in Child Care Center Licensing Regulations and Policies for 2014*. Fairfax, VA: Author. As of July 1, 2016: https://childcareta.acf.hhs.gov/sites/default/files/public/center_licensing_trends_brief_2014.pdf

National Center on Child Care Quality Improvement. (undated). *QRIS Resource Guide: Section 9, Consumer Education*, Washington, DC: Author. As of July 1, 2016: Office of Child Care, Administration for Children and Families, U.S. Department of Health and Human Services. As of July 1, 2016: https://qrisguide.acf.hhs.gov/files/chapters/QRISRG_Chapter_9_Consumer_Education.pdf

Norris, D. J., and L. Dunn. (2004). *Reaching for the Stars: Family Child Care Home Validation Study Final Report.* Stillwater and Norman, OK: Early Childhood Collaborative of Oklahoma. As of July 1, 2016: http://www.okdhs.org/OKDHS%20Report%20Library/ReachingForTheStarsFamilyChildCareValidationStudyFinalReport_dcc_05092007.pdf

Norris, D. J., L. Dunn, and L. Eckert. (2003). *Reaching for the Stars: Center Validation Study Final Report,* Stillwater, OK and Norman, OK: Early Childhood Collaborative of Oklahoma. As of July 1, 2016: http://www.okdhs.org/OKDHS%20Report%20Library/ReachingForTheStarsCenterValidationStudyFinalReport_dcc_05212007.pdf

Pianta, R. C., K. M. La Paro, and B. K. Hamre. (2008). *Classroom Assessment Scoring System (CLASS) Manual, Pre-K*. Baltimore, MD: Paul H. Brookes Publishing Company.

QRIS National Learning Network. (2016). *QRIS State Contacts and Map*. As of July 1, 2016: http://qrisnetwork.org/qris-state-contacts-map

Rainelli, S., R. Bulotsky-Shearer, V. Fernandez, D. Greenfield, and M. López. (2014). *Validity of the First Two Subtests of the PreLAS2000 as a Language Screener for Spanish Speaking Preschool Children*. Unpublished manuscript.

Sabol, T. J., S. S. Hong, R. C. Pianta, and M. R. Burchinal. (2013). Can rating pre-K programs predict children's learning? *Science* 34: 845–846.

Sabol, T. J., and R. C. Pianta. (2012). *Improving Child Care Quality: A Validation Study of the Virginia Star Quality Initiative*. Charlottesville, VA: University of Virginia Curry School of Education.

Schwartz, H. L., L. A. Karoly, V. Le, J. Tamargo, and C. M. Setodji. (2014). *Evaluation of Delaware Stars for Early Success: Year 1 Report*. RR-606-DOEL. Santa Monica, CA: RAND Corporation. As of July 1, 2016: http://www.rand.org/pubs/research_reports/RR606.html

Schilder, D., I. Iruka, H. Dichter, and D. Mathias. (2015). *Quality Rating and Improvement Systems: Stakeholder Theories of Change and Models of Practice*. As of July 1, 2016: http://qrisnetwork.org/sites/all/files/resources/2016-02-10%2009%3A21/QRIS%203.0%20Report%20V11%202016.2.5%20FINAL.pdf

Sirinides, P. (2010). *Demonstrating Quality: Pennsylvania Keystone STARS: 2010 Program Report*. Harrisburg, PA: Office of Child Development and Early Learning. As of July 1, 2016: http://www.pakeys.org/uploadedContent/Docs/Demonstrating%20Quality%202010.pdf

State of Delaware. (2011). *Race to the Top–Early Learning Challenge Application for Initial Funding*. As of July 1, 2016: https://www2.ed.gov/programs/racetothetop-earlylearningchallenge/applications/delaware.pdf

Swanson, K. (2013). *A Practical Outreach Guide for a Quality Rating and Improvement Systems (QRIS)*. Washington, D.C.: BUILD Initiative. As of July 1, 2016: http://qrisnetwork.org/sites/all/files/resources/mrobinson%40buildinitiative.org/2013-06-25%2013%3A26/A%20Practical%20Outreach%20Guide%20for%20a%20QRIS%20%28Clickable%29.pdf

Thornburg, K. R., W. A. Mayfield, J. S. Hawks, and K. L. Fuger. (2009). *The Missouri Quality Rating System School Readiness Study.* Columbia, MO: Center for Family Policy and Research. As of July 1, 2016: http://www.elcmdm.org/Knowledge%20Center/reports/MOQRSreport.pdf

Tout, K., R. Starr, T. Isner, J. Cleveland, M. Soli, and K. Quinn. (2010). *Evaluation of Parent Aware: Minnesota's Quality Rating System Pilot: Year 3 Evaluation Report*. Minneapolis, MN: Child Trends. As of July 1, 2016: https://s3.amazonaws.com/Omnera/VerV/s3finder/38/pdf/Parent_Aware_Year_3_Evaluation_Report_Nov_2010.pdf

Tout, K., R. Starr, L. Albertson-Junkans, M. Soli, and K. Quinn. (2011). *Evaluation of Parent Aware: Minnesota's Quality Rating System Pilot: Final Evaluation Report.* Minneapolis, MN: Child Trends. As of July 1, 2016: https://s3.amazonaws.com/Omnera/VerV/s3finder/38/pdf/Parent_Aware_Year_4_Final_Evaluation_Technical_Report_Dec_2011.pdf

Tout, K., J. Cleveland, W. Li, R. Starr, M. Soli, and E. Bultinck. (2016). *Parent Aware: Minnesota's Quality Rating and Improvement System Initial Validation Report*. Minneapolis, MN: Child Trends. As of July 1, 2016: http://www.childtrends.org/wp-content/uploads/2016/03/Parent-Aware-Validation-and-Letter-3-2016.pdf

Vitiello, V. E. (2014). Using the CLASS measure in family child care homes. Charlottesville, VA: Teachstone. As of July 1, 2016: http://www.ccfc.ca.gov/pdf/programs/impact/training/AMCH4_CLASS%20Observer_Trainer%20Tools-Resources/4_2_4CLASS-FCC-White-Paper.pdf

Vogel, C., N. Aikens, S. Atkins-Burnett, S., E. S. Martin, M. Caspe, S. Sprachman, S., and J. M. Love. (2008). *Reliability and Validity of Child Outcome Measures with Culturally and Linguistically Diverse Preschoolers: The First 5 LA Universal Preschool Child Outcomes Study Spring 2007 Pilot Study*. As of July 1, 2016: http://es.www.first5la.org/files/UPCOSTechMemoSpringPilotReportFINAL.pdf

Weiland, C., K. Ulvestad, J. Sachs, and H. Yoshikawa. (2013). Associations between classroom quality and children's vocabulary and executive function skills in an urban public prekindergarten program. *Early Childhood Research Quarterly 28* (2): 199–209.

Weiland, C., and H. Yoshikawa. (2013). Impacts of prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. *Child Development,* 84(6): 2112–2130.

Woodcock, R. W., K. McGrew, and N. Mather. (2006). *Woodcock-Johnson III, Normative Update, Tests of Achievement*. Itasca, IL: Riverside Publishing.

Woodcock, R. W., A. F. Muñoz-Sandoval, K. McGrew, N. Mather, and F. Schrank. (2007). *Batería III, Normative Update, Woodcock-Muñoz*. Itasca, IL: Riverside Publishing.

Zaslow, M., and K. Tout. (2014). *Reviewing and Clarifying Goals, Outcomes and Levels of Implementation: Toward the Next Generation of Quality Rating and Improvement Systems (QRIS).* OPRE Research Brief #2014-75. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. As of July 1, 2016: http://www.acf.hhs.gov/sites/default/files/opre/qris_conceptual_framework.pdf

Zellman, G. L., and R. Fiene. (2012). *Validation of Quality Rating and Improvement Systems for Early Care and Education and School-Age Care.* Research-to-Policy, Research-to-Practice Brief. OPRE 2012-29. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Zellman, G. L., and M. Perlman. (2008). *Child-Care Quality Rating and Improvement Systems in Five Pioneer States: Implementation Issues and Lessons Learned*. MG-795-AECF/SPF/UWA. Santa Monica, CA: RAND Corporation. As of July 1, 2016: http://www.rand.org/pubs/monographs/MG795.html

Zellman, G. L., M. Perlman, V. Le, and C. M. Setodji. (2008). *Assessing the Validity of the Qualistar Early Learning Quality Rating and Improvement System as a Tool for Improving Child-Care Quality.* MG-650-QEL. Santa Monica, CA: RAND Corporation. As of July 1, 2016: http://www.rand.org/pubs/monographs/MG650.html